Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.51, p. 28-34. January/February., 2025.

DOI: https://doi.org/10.5935/jetia.v11i51.1367

RESEARCH ARTICLE OPEN ACCESS

CLASSIFICATION OF PROMINENT CACAO POD DISEASES USING MULTI-FEATURE VISUAL ANALYSIS AND K-NEAREST NEIGHBORS ALGORITHM

Earl Clarence S. San Diego[1], Seph Gerald C. Rodrin[2], Edwin R. Arboleda[3]

[1,2,3] Department of Computer, Electronics and Electrical Engineering, Cavite State University, Don Severino Delas Alas Campus, Indang, Cavite, Philippines

[1]http://orcid.org/0009-0007-6766-0205, [2]http://orcid.org/0009-0009-6724-1240, [3]http://orcid.org/0000-0001-9371-8895

Email: [1]main.earlclarence.sandiego@cvsu.edu.ph, [2]main.sephgerald.rodrin@cvsu.edu.ph, 3edwin.r.arboleda@cvsu.edu.ph

ARTICLE INFO

Article History
Received: November 28, 2024
Revised: December 20, 2024
Accepted: January 15, 2025
Published: January 30, 2025

Keywords:
Cacao Pod,
Visual Feature,
Feature Extraction,
k-Nearest Neighbors.

ABSTRACT

Cacao has been one of the most promising crops produced in the Philippines due to its increasing demand in various local and international markets. Although cacao production aspired to be heightened to cope with the global trend, several difficulties were still needed to be addressed in crop propagation, mainly due to disruptive diseases and pests. In response to this problem, the study devised an algorithm based on k-Nearest Neighbors that can detect whenever a cacao pod was infected with the three most prominent diseases: black pod rot, Monilia, and pod borer infestations. The machine training model was preceded with visual feature extraction of color and texture parameters representing the cacao pod samples. It was found that the fine k-Nearest Neighbors algorithm achieved the highest validation and testing accuracies of 93.44% and 96.67%, respectively. The study's outcome suggested the continuous practicality of fusing visual feature extraction processes with supervised machine learning to generate models that can be applied to improve agricultural methods.

I. INTRODUCTION

Due to international demand and competitive industry, cacao has been included as one of the crucial crops in the strategic and technology plan for agricultural resources research in the Philippines. It has been recommended that the commercial production of cacao be boosted to accommodate the rising global demand and take advantage of the country's favorable geographic landscape. Although efforts and initiatives are present, the cocoa business in the Philippines still suffers from the risks presented by possible attacks of pests and diseases [1].

The black pod rot disease, caused by Phytophthora species, and the frosty pod rot, also called the Monilia disease, were the most prominent cacao pod ailments worldwide [2]. Regarding pests, the pod borer, Conopomorpha cramerella, significantly contributed to the disruptions in cacao production in Southeast Asia [3]. Small-scale local farmers were projected to face more damaging consequences due to lack of agricultural knowledge and skills in planting and post-harvesting management of cacao crops. Due to the devastations brought by emerging diseases, cacao plantations were expected to decline without appropriate intervention, contributing to significant production losses. In response to these conflicts, national agencies and industry partners have developed programs for disease management and effective resource use. These activities aim to utilize cutting-edge technologies and tools to promote long-term responses for sustainable cacao health and industry firmness over time [2].

Artificial intelligence-based techniques, like deep learning and machine learning, have been infamous for developing programs meant to diagnose plant diseases at their initial stages, paving the way for early preventive maintenance and actions. When complemented by high-resolution photography, these algorithms will break free from the restrictions attributed to the manual classification of diseased crops, often influenced by subjective observations [4].

Neural networks, such as convolutional neural networks (CNNs), and supervised machine learning models were typically applied to tools and gadgets that can perceive the presence of plant defects with high classification accuracy. These approaches heavily depended on data, such as standard pictures or sensor data, as the number of samples significantly shapes their detection capabilities [5]. For traditional machine learning methods, image processing techniques were preferably embedded for gathering necessary visual features, like shape, color, and texture, which can later be administered to classification algorithms for training.

On the other hand, neural networks can facilitate both feature extraction and classification training processes, making them suitable for more extensive disease differentiation [6]. This ability of the neural networks was showcased in the studies of [7-11] as applied to actual cacao pods and leaves.

Brought by widespread endeavors to combat the implications of unsolicited cacao threats, functional mobile applications have been launched to identify several cacao pests and diseases. [12] created AuToDiDAC, a mobile application aimed at detecting and assessing the level of black pod rot infection in cacao pods. With the application of graph cuts, color balancing, and fast k-means clustering algorithms for processing raw cacao pod images, the SVM classifier incorporated in the tool was found to have a classification accuracy of 84% when evaluated in ten independent pod samples.

Furthermore, [13] designed an image processing application that can identify general initial symptoms of pest and disease infestation in cacao, achieving an accuracy of 100%. The local binary pattern (LBP) and Gabor filter algorithm were utilized for feature extraction, while deep learning methods were used for the classifier.

The resulting mobile application was programmed to use cellphone camera captures as input data. On the other hand, [14] has integrated modified CNN to identify pods infected with swollen shoot disease among cacao trees. According to [15] worked with a CNN-based smartphone application named Cocoa Companion that can detect swollen shoots as well with the addition of black pod diseases, having a maximum accuracy of 80%. SSD MobileNet V2 was selected among the other three options for the CNN architecture: EfficientDetD0, CenterNet ResNet50 V2, and SSD ResNet50 V1 FPN. According to [16] improved the former study by considering a higher sample of cacao pods for training, making the accuracy as high as 88%.

Aside from constructing actual image processing programs, various researchers have chosen to be engaged in finding the perfect combination of feature extraction techniques, supervised machine learning algorithms, and neural networks to form a model that can accurately determine if a cacao pod is healthy or infected by distinct diseases.

Several studies have used conventional image processing techniques to obtain relevant visual information from cacao pod images, then utilized algorithms such as SVM and neural networks for the classification model. According to [17] employed the Haralick algorithm to extract texture features in cacao pods, then used the parameters for the classification training of six machine learning algorithms: Naïve Bayes, Decision Stump, Random Forest, Hoeffding Tree, Multilayer Neural Network, and CNN. CNN achieved 99% accuracy, the highest among the six, in recognizing the Phytophthora palmivora disease or black pod rot infection in cacao pods. Another study of [18] worked with the normalization of RGB (Red, Green, Blue) parameters of cacao pod images to transform them into hue, saturation, and value (HSV) features.

These were then placed into an untrained k-Nearest Neighbors (KNN) classifier to distinguish three class categories: fruit rot disease, fruit-sucking ladybugs, and pod pests. After conducting k-fold cross-validation, the best accuracy attained is 99.33%, at a k-value of 5. According to, [19] created a classification model for detecting if a cacao pod is diseased using a histogram of oriented gradients (HoG) and local binary pattern (LBP) for feature extraction.

Three classification algorithms were trained: SVM, random forest, and artificial neural network (ANN), with ANN attaining the highest classification accuracy of 85%. The study of [20] performed the same procedure with slight modifications of the utilized methods. Color histogram was used instead of HoG, and random forest was replaced by logistic regression (LR). Still, ANN gained the highest classification accuracy of 98.3%.

Pre-existing studies have mostly attempted to devise models for detecting two particular cacao pod diseases. For instance, [21] used five CNN architectures: custom CNN, VGG16, EfficientNetB0, Resnet50, and LeNet-5, as both visual extraction and classification systems for diagnosis of black pod rot or pod borer disease, achieving a maximum accuracy of 91.79%. Conversely, [22] utilized other CNN architectures, with EfficientNetB0 garnering a higher maximum accuracy of 94%. [23] used CNN and the stochastic gradient descent (SGD) algorithm to detect black pod rot and the mistletoe disease. Moreover, [24-26] worked with the identification of the Phytophthora and Monilia diseases.

With these studies as references, a gap can be found in the versatility of the pre-established models in identifying more than two types of disease and pest infestation in cacao pods.

Moreover, image processing techniques in feature analysis were gradually overlooked due to the emergence of different CNN architectures that can facilitate feature extraction and classification. With that, this study utilized conventional visual feature extraction processes to extract RGB, HSV, and gray-level co-occurrence matrix (GLCM) texture parameters from cacao pod samples. Those values were used to model a KNN algorithm for detecting the presence of black pod rot, Monilia, or pod borer disease in cacao pods.

Specifically, it was aimed to: (1) extract numerical parameters representing the RGB, HSV, and texture features of cacao pod images, (2) train a classification model based on three KNN types: fine, cosine, and weighted KNN, and (3) evaluate the KNN models in terms of their accuracy, precision, and recall in identifying specified cacao pod diseases. With the objectives being met, a flexible machine learning model was generated to detect the three most commonly observed illnesses of cacao pods.

II. MATERIALS AND METHODS

This section presents the research methodology employed in the study to classify cacao pod diseases using multi-feature visual analysis and k-nearest neighbor's algorithm. Specifically, MATLAB software was used for performing feature extraction and disease classification.

II.1 MATERIALS

The data used for this study consisted of 800 images of both diseased and healthy cacao pods that were readily obtained from Kaggle. Specifically, a total of 200 images per class for four distinct cacao pod classifications were utilized.

Namely, they are: healthy, infected with black pod rot, infected with pod borers, and Monilia-diseased. Figure 1 presents the sample images of the cacao pods used for the study. To ensure an unbiased model evaluation, the dataset was split into three subsets: training (70%), validation (15%), and testing (15%).

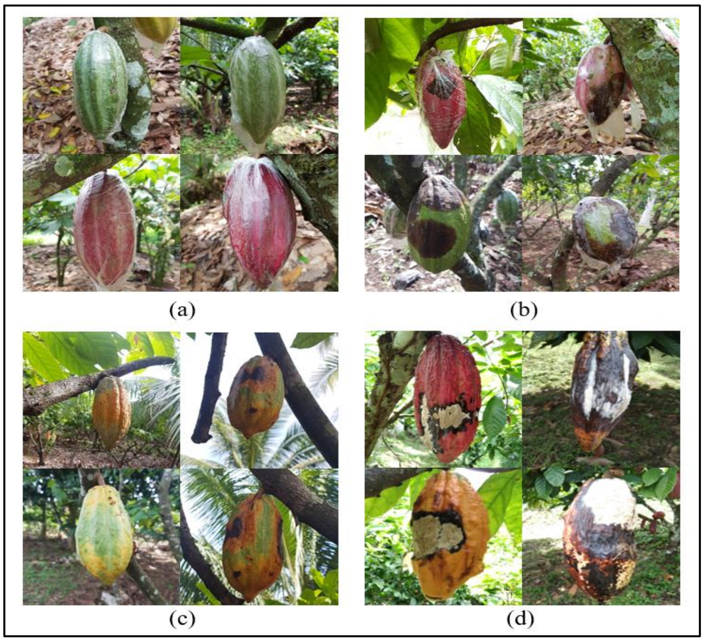

Figure 1: Sample raw cacao pod images: (a) healthy, (b) infected with black pod rot, (c) infested with pod borers, and (d) Monilia-diseased.
Source: Authors, (2025).

II.2 METHODS

Figure 2 shows the overall procedural framework for the study. It consists of four major steps that were followed methodically to ensure proper data acquisition. Specifically, the steps include: (1) image acquisition and pre-processing, (2) visual feature extraction, (3) KNN classification model training, and (4) final testing and evaluation of the model.

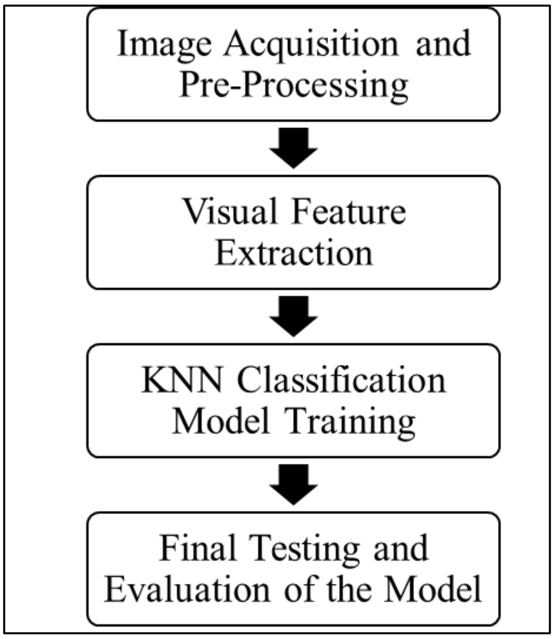

Figure 2: Procedural framework for the study.
Source: Authors, (2025).

II.2.1 IMAGE ACQUISITION AND PRE-PROCESSING

Image pre-processing was done to filter out the data and retain only good-quality images that will be used for the disease classification. Images that were blurry and those that contained multiple diseases were excluded from the data. After that, the filtered images were subjected to pre-processing through background subtraction to separate the target foreground object from the background. Lastly, the resulting images were subjected to ROI (region of interest) selection to ensure that the most significant features of the images were retained and irrelevant information that may affect the classification process was removed. The pre-processing procedures administered were visualized as shown below in Figure 3.

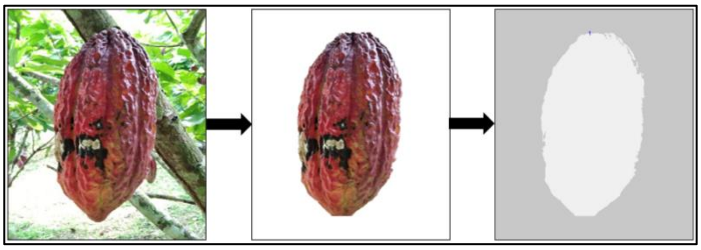
Figure 3: Pre-processing process involving background subtraction and ROI selection.
Source: Authors, (2025).

II.2.2 VISUAL FEATURE EXTRACTION

The visual feature extraction process utilized a combination of color and texture characteristics for a more comprehensive representation of the visual data needed for KNN classification. Color features were extracted by computing the means of the red, green, and blue channels to represent the color distribution of the various images. Similarly, the mean of the hue, saturation, and value channels were also computed to depict the images' perceptual attributes. Figure 4 and Figure 5 show a snippet of the code used to extract the color features.

```
% Compute mean RGB values
meanR = mean(mean(img(:,:,1)));
meanG = mean(mean(img(:,:,2)));
meanB = mean(mean(img(:,:,3)));
```

Figure 4: Snippet of code used for extracting RGB values.
Source: Authors, (2025).

```
% Convert RGB to HSV
imgHSV = rgb2hsv(img);

% Compute mean HSV values
meanH = mean(mean(imgHSV(:,:,1)));
meanS = mean(mean(imgHSV(:,:,2)));
meanV = mean(mean(imgHSV(:,:,3)));
```

Figure 5: Snippet of code used for extracting HSV values.
Source: Authors, (2025).

On the other hand, to characterize the texture of the cacao pods, GLCM was used to extract the following features: energy, entropy, homogeneity, and contrast. These features were extracted to capture and give numerical values to the irregularities and patterns that occur due to cacao pod infections. Figure 6 shows the code used to extract the GLCM-based texture values.

```
% Compute the Gray-Level Co-Occurrence Matrix (GLCM)
glcm = graycomatrix(img, 'Offset', [0 1; -1 1; -1 0; -1 -1], ...
    'Symmetric', true);
stats = graycoprops(glcm, {'Contrast', 'Energy', ...
    'Homogeneity'});

% Compute entropy (from normalized GLCM)
glcmNorm = glcm ./ sum(glcm, 'all');
entropyVal = -sum(glcmNorm(:) .* log2(glcmNorm(:) + eps));
```

Figure 6: Snippet of code used for extracting GLCM-based texture values.
Source: Authors, (2024).

II.2.3 KNN CLASSIFICATION MODEL TRAINING

Following the visual feature extraction of the RGB, HSV, and GLCM-based texture values, the KNN machine learning algorithm was applied to classify the cacao pod diseases. Specifically, the MATLAB's Classification Learner App was used to classify the diseases using three KNN variants, namely: fine KNN, cosine KNN, and weighted KNN. These three variants were used since each offers different metrics and weighting strategies that optimize the classification of the dataset. Lastly, holdout validation was implemented as the validation process to ensure the model's accuracy and generalization capabilities.

II.2.4 FINAL TESTING AND EVALUATION OF THE MODEL

The study employed a dataset splitting of 70-15-15 percent for training, validation, and testing to classify the various cacao pod diseases. Specifically, accuracy, precision, and recall were utilized to assess the performance of the three KNN variants.

Accuracy is defined as the ratio of correctly predicted observations to the total number of observations. It is ideal for symmetric data sets that exhibit virtually equal false positive and false negative values. It quantifies the overall accuracy of the model's classification.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \qquad (1)$$

Precision denotes the accuracy of the positive predictions produced by the mode. It is determined by dividing the total number of data points accurately classified by the model by the number of true positives.

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

The true positive rate, or recall, assesses the classifier's ability to correctly identify all actual positive instances. The calculation involved dividing the overall count of positive data points by the count of actual positive data points.

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

III. RESULTS AND DISCUSSIONS

This section presents the findings and results gathered from the employed methodologies, starting from the visual feature extraction and the cacao pod diseases classification using the KNN algorithm.

III.1 EXTRACTION OF VISUAL FEATURES

Before classifying the various cacao pod diseases, the visual features of each raw cacao pod image were first extracted using the MATLAB code shown in Figures 4 to 6. Running the code facilitated the feature extraction by transforming the image features, such as RGB, HSV, and textures, to numerical representations stored in a table. Figure 7 shows the sample RGB features extracted from the samples. It consists of five columns, with the first column containing the image file name and the second to the fourth column comprising the calculated mean values for red, green, and blue features, respectively. Meanwhile, the last column indicates the class or specific disease of the images. These mean values of each feature served as the predictor for establishing the distinctions between the cacao pod classes to distinguish them from one another.

1	ImageName	MeanRed	MeanGreen	MeanBlue	Class
2	blackrot_1.jpg	174.92021	169.985444	166.514708	black_rot
3	blackrot_10.jpg	183.501442	177.102083	130.464263	black_rot
4	blackrot_100.jpg	173.672149	174.843498	158.163681	black_rot
5	blackrot_101.jpg	161.535107	155.974533	158.395814	black_rot
6	blackrot_102.jpg	162.92057	163.94649	143.337524	black_rot
7	blackrot_103.jpg	176.20407	168.867925	158.828487	black_rot
8	blackrot_104.jpg	213.362148	203.584686	168.608157	black_rot
9	blackrot_105.jpg	158.813604	152.971441	148.975583	black_rot
10	blackrot_106.jpg	158.813604	152.971441	148.975583	black_rot
11	blackrot_107.jpg	140.832701	138.565519	135.098278	black_rot
12	blackrot_108.jpg	156.688889	145.839046	144.096673	black_rot
13	blackrot_109.jpg	142.924948	142.399286	129.63372	black_rot
14	blackrot_11.jpg	194.787386	190.761204	190.198213	black_rot
15	blackrot_110.jpg	162.46537	163.494382	142.781598	black_rot
16	blackrot_111.jpg	168.54512	157.216695	161.923801	black_rot
17	blackrot_112.jpg	155.758307	144.350208	124.574117	black_rot
18	blackrot_113.jpg	143.471842	145.613282	133.908678	black_rot
19	blackrot_114.jpg	147.438755	157.166834	154.691432	black_rot
20	blackrot_115.jpg	162.972009	153.58911	147.791718	black_rot

Figure 7: Sample RGB values extracted from the samples.
Source: Authors, (2024).

Consequently, Figure 8 displays the sample HSV values extracted from the different classes. Like the extracted RGB features, it also consists of five columns, with the file name of the images displayed on the first one. Meanwhile, the second, third, and fourth columns display the mean values extracted for hue, saturation, and value, respectively. The class of the images was also indicated in the last column. These extracted features were also used as cacao pod disease classification parameters.

1	ImageName	MeanHue	MeanSaturation	MeanValue	Class
2	blackrot_1.jpg	0.197938928	0.114234942	0.689639611	black_rot
3	blackrot_10.jpg	0.107698812	0.338908974	0.724267848	black_rot
4	blackrot_100.jpg	0.17544311	0.165497944	0.703903077	black_rot
5	blackrot_101.jpg	0.390220793	0.132974184	0.647231876	black_rot
6	blackrot_102.jpg	0.126252445	0.270481735	0.653127243	black_rot
7	blackrot_103.jpg	0.151402373	0.167769806	0.701004693	black_rot
8	blackrot_104.jpg	0.164745506	0.237659545	0.841638284	black_rot
9	blackrot_105.jpg	0.144121615	0.117808064	0.627371472	black_rot
10	blackrot_106.jpg	0.144121615	0.117808064	0.627371472	black_rot
11	blackrot_107.jpg	0.187707535	0.109732976	0.556268473	black_rot
12	blackrot_108.jpg	0.187994312	0.192420358	0.618981574	black_rot
13	blackrot_109.jpg	0.247690075	0.208943933	0.577631783	black_rot
14	blackrot_11.jpg	0.231028464	0.082446862	0.773352902	black_rot
15	blackrot_110.jpg	0.126854594	0.271850769	0.651420141	black_rot
16	blackrot_111.jpg	0.395104364	0.143912838	0.669795192	black_rot
17	blackrot_112.jpg	0.077171238	0.335072503	0.61324293	black_rot
18	blackrot_113.jpg	0.202955389	0.207852113	0.589114522	black_rot
19	blackrot_114.jpg	0.282917813	0.115368145	0.625218521	black_rot
20	blackrot_115.jpg	0.153080504	0.180845712	0.644418982	black_rot

Figure 8: Sample HSV values extracted from the samples.
Source: Authors, (2024).

The last set of codes extracted the texture features of the images. The sample extracted values for each of them are displayed in Figure 9. Column one indicates the specific file name of the images, and columns two to five shows the values extracted for the various texture parameters, namely entropy, contrast, energy, and homogeneity. Like the previous features, the last column also indicates the class or specific disease of the images. The numerical representations of the texture extracted from the images enhanced the classification capabilities of the machine learning algorithm by providing more parameters to be fed into the system.

1	ImageName	Entropy	Contrast	Energy	Homogeneity	Class
2	blackrot_1.jpg	5.88264277	0.627607162	0.142402569	0.836931417	black_rot
3	blackrot_10.jpg	5.22315327	0.498099549	0.189491263	0.889756439	black_rot
4	blackrot_100.jpg	5.59336857	0.332917392	0.132674474	0.870545011	black_rot
5	blackrot_101.jpg	5.73904725	0.860616204	0.168875598	0.825329384	black_rot
6	blackrot_102.jpg	5.46394993	0.415148938	0.186113124	0.875376876	black_rot
7	blackrot_103.jpg	5.87713697	0.625541091	0.126421237	0.832169895	black_rot
8	blackrot_104.jpg	5.21568838	0.489077708	0.231795935	0.87438798	black_rot
9	blackrot_105.jpg	5.91299084	0.70921349	0.118900637	0.816637672	black_rot
10	blackrot_106.jpg	5.90856447	0.697451577	0.119176487	0.817153582	black_rot
11	blackrot_107.jpg	5.55044541	0.514322465	0.156201476	0.864463907	black_rot
12	blackrot_108.jpg	5.47662709	0.709734754	0.191581369	0.852680039	black_rot
13	blackrot_109.jpg	5.6511286	0.522071479	0.144276714	0.873185094	black_rot
14	blackrot_11.jpg	5.61144635	0.623425025	0.195588801	0.838544832	black_rot
15	blackrot_110.jpg	5.4737132	0.417277592	0.184152005	0.874862191	black_rot
16	blackrot_111.jpg	5.94956508	0.811224638	0.149636812	0.817202128	black_rot
17	blackrot_112.jpg	5.59322993	0.697644324	0.162633324	0.862538672	black_rot
18	blackrot_113.jpg	5.93341899	0.461326822	0.108322308	0.869319454	black_rot
19	blackrot_114.jpg	5.88644016	0.522268194	0.125385091	0.842533089	black_rot
20	blackrot_115.jpg	5.92623405	0.704857711	0.132831633	0.82970965	black_rot

Figure 9: Sample GLCM-based texture values extracted from the samples.
Source: Authors, (2025).

Table 1: Summary of visual features extracted from the cacao pods.

Visual Features	**Cacao Pod Class**			
	Healthy	**Black Pod Rot**	**Monilia**	**Pod Borer**
Mean Red	133.468 - 209.654	124.347 - 213.368	108.833 - 224.001	125.189 - 205.783
Mean Green	144.545 - 212.717	124.067 - 207.290	82.547 - 227.259	117.508 - 208.939
Mean Blue	120.550 - 192.781	106.428 - 190.198	73.471 - 225.346	90.109 - 178.892
Mean Hue	0.123 - 0.310	0.077 – 0.510	0.102 - 0.730	0.062 - 0.226
Mean Saturation	0.105 - 0.380	0.074 – 0.341	0.036 - 0.524	0.129 - 0.496
Mean Value	0.568 - 0.841	0.504 – 0.842	0.445 - 0.897	0.495 - 0.838
Entropy	4.713 - 5.716	4.849 - 6.176	4.314 - 6.261	4.331 - 5.774
Contrast	0.216 - 0.531	0.312 - 0.969	0.304 - 1.173	0.234 - 0.668
Energy	0.118 - 0.258	0.100 - 0.306	0.083 - 0.479	0.114 - 0.335
Homogeneity	0.848 - 0.914	0.787 - 0.914	0.778 - 0.904	0.828 - 0.927

Source: Authors, (2025).

Table 1 contains a summary of the visual features extracted for every cacao pod class. These numerical parameters served as inputs for the KNN classification training, specifically using three types of KNN: fine KNN, cosine KNN, and weighted KNN.

III.2 KNN CLASSIFICATION MODEL TRAINING

After acquiring the numerical parameters representing cacao pods' RGB, HSV, and texture features, the KNN machine learning algorithm was trained using MATLAB’s Classification Learner App. The classification session was driven by the predictors obtained from the previous procedure. Moreover, a holdout validation scheme was used by setting aside 15% of the original sample for the preliminary assessment of the KNN algorithm. A total of 15% of the dataset was also removed to assess the performance of the KNN classifiers in classifying cacao pod diseases.

For the KNN training, the study utilized three models, namely: fine KNN, cosine KNN, and weighted KNN. This is to provide insights regarding which model classifies the cacao pod diseases more accurately.

Each of their figure of merits was obtained after subjecting the extracted features to the various KNN training models. Specifically, their validation confusion matrices were consolidated to collect the numerical parameters needed to calculate the other figure of merits. Table 2 summarizes the results of the model evaluation in terms of accuracy, precision, and recall.

Table 2: Validation results for the KNN classification models.

Model	Accuracy	Precision	Recall
Fine KNN	93.44%	93.57%	93.41%
Cosine KNN	68.85%	67.69%	68.68%
Weighted KNN	91.80%	91.99%	91.85%

Source: Authors, (2025).

Table 2 shows that the fine KNN model performed the best in classifying various cacao pod diseases among all other models during the validation phase. It showed accuracy, precision, and recall of 93.44%, 93.57%, and 93.41%, respectively. To further elaborate, the corresponding confusion matrix for the Fine KNN model is shown in Figure 10. Out of 30 cacao pods with black pod rot disease, 25 were classified correctly, one was misclassified as healthy, and four were misclassified as being infected by Monilia disease. Likewise, for those 31 healthy cacao pods, 29 were correctly labeled, one was misclassified as having black pod rot, and one was misclassified as being infected with pod borers. Among 31 monilia-diseased pods, 30 were correctly identified, and only one was miscategorized as having black pod rot. Lastly, all 30 pods infested with pod borers were correctly categorized by the fine KNN algorithm.

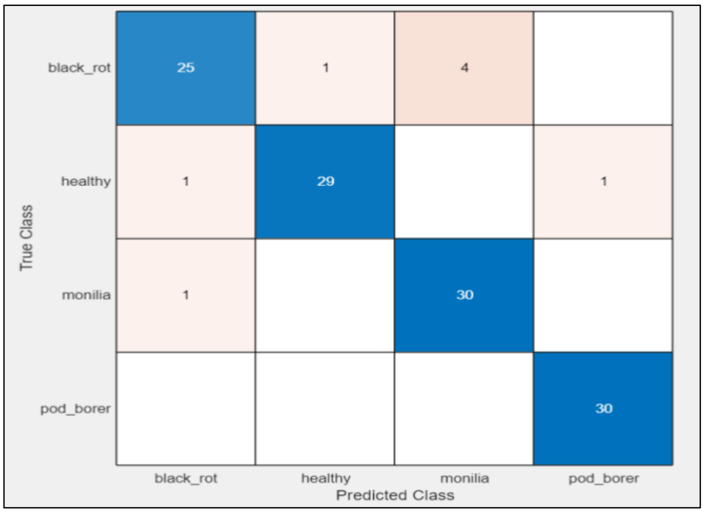

Figure 10: Confusion matrix for the validation results of the fine KNN algorithm.
Source: Authors, (2025).

III.3 EVALUATION OF THE PROPOSED CLASSIFICATION MODEL

In addition to the initial evaluation of the KNN machine learning algorithm model's performance, the holdout validation method assisted the further optimization of the algorithms to adapt better and adjust to new data. Upon finalizing the classification models, each was independently subjected to testing using the remaining samples. Like the holdout validation, the accuracy, precision, and recall of the KNN models were assessed in the final performance evaluation. Table 3 shows the summary of the evaluation of the results.

Table 3: Evaluation results for the KNN classification models.

Model	Accuracy	Precision	Recall
Fine KNN	96.67%	96.67%	96.67%
Cosine KNN	78.33%	77.84%	78.33%
Weighted KNN	95.83%	95.86%	95.83%

Source: Authors, (2025).

Similar to the validation findings, the fine KNN model performed best in classifying the different cacao pod diseases. Table 3 shows that it now has an accuracy, precision, and recall of 96.67% for all three figures of merits. The confusion matrix provided in Figure 11 shows the correctness of the model in classifying each class. For 30 cacao pods with black pod rot, 28 were correctly classified, and two were mislabeled as infected with monilia disease. Among the 30 healthy pods, all of them were classified correctly. Meanwhile, for those that are infected with monilia disease, 28 were correctly labeled, and two were miscategorized as having black pod rot. Finally, all 30 pods with pod borer disease were correctly labeled by the fine KNN classifier.

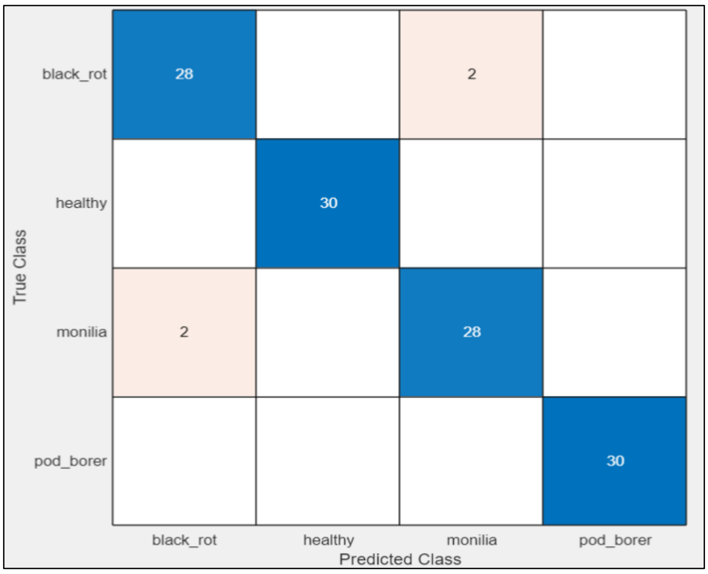

Figure 11: Confusion matrix for the evaluation results of the fine KNN algorithm.
Source: Authors, (2024).

IV. CONCLUSIONS

The present study focused on identifying prominent cacao pod diseases through multi-feature visual analysis combined with the KNN machine learning algorithm. Specifically, the RGB, HSV, and GLCM-based texture features were considered for the visual analysis of 800 cacao pod images that were classified into four classes: healthy, black pod rot-infected, pod borer-infested, and monilia-diseased. The numerical representations of each feature were then fed as predictors for the training of the KNN classifier using three models evaluated to identify the most optimum result. From the three KNN models, it was revealed that fine KNN achieved the highest accuracy for both the validation and testing stages, recording 93.44% and 96.67%, respectively. These results highlight the effectiveness and reliability of combining multi-feature visual analysis and KNN algorithms to distinguish between cacao pod diseases. This approach provides a valuable contribution to agriculture, especially in cacao disease management, as a tool for early disease detection and monitoring. For further improvements in the research, future researchers may add other cacao pod diseases and pests, such as swollen shoots, to further expand the diagnosing capabilities of the model. Likewise, other KNN models or variants not used in the study may be evaluated for their potential to enhance the accuracy of the overall classification scheme. In addition, additional relevant cacao-related applications may be explored, including cacao bean grading and quality assessment, whereby the combination of multi-feature visual extraction and KNN algorithms can be implemented.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Methodology: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Investigation: Earl Clarence S. San Diego and Edwin R. Arboleda.
Discussion of results: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Writing – Original Draft: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Writing – Review and Editing: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Resources: Earl Clarence S. San Diego and Seph Gerald C. Rodrin.
Supervision: Earl Clarence S. San Diego, Seph Gerald C. Rodrin, and Edwin R. Arboleda.
Approval of the final text: Earl Clarence S. San Diego, Seph Gerald C. Rodrin, Edwin R. Arboleda.

VI. ACKNOWLEDGMENTS

The authors want to express their sincerest gratitude and appreciation to the department, the institution, and other relevant individuals who made the conduct of the study possible.

VII. REFERENCES

[1] W. N. Magallon, M. E. Patalinghug, and M. G. G. Tangalin, "Status of Cacao (Theobroma cacao L.) production on its challenges and prospect in Zamboanga del Norte Province in the Philippines," International Journal of Agricultural Technology, vol. 18, no. 3, 2022.

[2] J. P. Marelli et al., "Chocolate under threat from old and new cacao diseases," Phytopathology, vol. 109, no. 8. 2019. doi: 10.1094/PHYTO-12-18-0477-RVW.

[3] J. Niogret et al., "Host preferences of the cocoa pod borer, Conopomorpha cramerella, the main threat to cocoa production in Southeast Asia," Entomologia Experimentalis et Applicata, vol. 168, no. 3, 2020, doi: 10.1111/eea.12882.

[4] I. Ahmed and P. K. Yadav, "Plant disease detection using machine learning approaches," Expert Systems, vol. 40, no. 5, 2023, doi: 10.1111/exsy.13136.

[5] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," Microprocessors and Microsystems, vol. 80, 2021, doi: 10.1016/j.micpro.2020.103615.

[6] M. Shoaib et al., "An advanced deep learning models-based plant disease detection: A review of recent research," Frontiers in Plant Science, vol. 14. 2023. doi: 10.3389/fpls.2023.1158933.

[7] A. F. M. Harvyanti, R. I. Baihaki, Dafik, Z. R. Ridlo, and I. H. Agustin, "Application of Convolutional Neural Network for Identifying Cocoa Leaf Disease," 2023. doi: 10.2991/978-94-6463-174-6_21.

[8] J. Atuhurra, Y.-R. D. N'guessan, and L. Pabitra, "Image Classification for CSSVD Detection in Cacao Plants." [Online]. Available: https://arxiv.org/pdf/2405.04535

[9] D. O. Cagadas and R. A. Labajan, "Leaf-Based Cacao Diseases Classification Using Image Processing," Sci.Int.(Lahore), vol. 35, no. 4, pp. 369–373, 2024.

[10] R. A. Godmalin, C. J. Aliac, and L. Feliscuzo, "Cacao Pod Infection Level Classification Using Transfer Learning," in 2023 IEEE Open Conference of Electrical, Electronic and Information Sciences, eStream 2023 - Proceedings, 2023. doi: 10.1109/eStream59056.2023.10135062.

[11] R. Y. Montesino, J. A. Rosales-Huamani, and J. L. Castillo-Sequera, "Detection of phytophthora palmivora in cocoa fruit with deep learning," in Iberian Conference on Information Systems and Technologies, CISTI, 2021. doi: 10.23919/CISTI52073.2021.9476279.

[12] D. S. Tan et al., "AuToDiDAC: Automated Tool for Disease Detection and Assessment for Cacao Black Pod Rot," Crop Protection, vol. 103, 2018, doi: 10.1016/j.cropro.2017.09.017.

[13] Basri, R. Tamin, H. A. Karim, Indrabayu, and I. S. Areni, "Mobile image processing application for cacao's fruits pest and disease attack using deep learning algorithm," ICIC Express Letters, vol. 14, no. 10, 2020, doi: 10.24507/icicel.14.10.1025.

[14] M. Coulibaly, K. H. Kouassi, S. Kolo, and O. Asseu, "Detection of 'Swollen Shoot' Disease in Ivorian Cocoa Trees via Convolutional Neural Networks," Engineering, vol. 12, no. 03, 2020, doi: 10.4236/eng.2020.123014.

[15] S. Kumi, D. Kelly, J. Woodstuff, R. K. Lomotey, R. Orji, and R. Deters, "Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection," in Procedia Computer Science, 2022. doi: 10.1016/j.procs.2022.07.013.

[16] R. K. Lomotey, S. Kumi, R. Orji, and R. Deters, "Automatic detection and diagnosis of cocoa diseases using mobile tech and deep learning," International Journal of Sustainable Agricultural Management and Informatics, vol. 10, no. 1, 2024, doi: 10.1504/IJSAMI.2024.135403.

[17] J. B. Rola et al., "Convolutional Neural Network Model for Cacao Phytophthora Palmivora Disease Recognition," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 15, no. 8. 2024. [Online]. Available: www.ijacsa.thesai.org

[18] M. Y. Pusadan, Syahrullah, Merry, and A. I. Abdullah, "k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 6, no. 3, 2022, doi: 10.29207/resti.v6i3.4064.

[19] C. Rodriguez, O. Alfaro, P. Paredes, D. Esenarro, and F. Hilario, "Machine Learning Techniques in the Detection of Cocoa (Theobroma cacao L.) Diseases," Annalsofrscb.Ro, vol. 25, no. 3, 2021.

[20] N. G. Baculio and J. B. Barbosa, "An Objective Classification Approach of Cacao Pods using Local Binary Pattern Features and Artificial Neural Network Architecture (ANN)," Indian Journal of Science and Technology, vol. 15, no. 11, pp. 495–504, Mar. 2022, doi: 10.17485/IJST/v15i11.60.

[21] K. S. Soh, E. G. Moung, K. J. J. Danker, J. A. Dargham, and A. Farzamnia, "Cocoa Diseases Classification using Deep Learning Algorithm," ITM Web of Conferences, vol. 63, 2024, doi: 10.1051/itmconf/20246301014.

[22] R. A. Godmalin, C. J. Aliac, and L. Feliscuzo, "Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm," in 4th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2022, 2022. doi: 10.1109/IICAIET55139.2022.9936817.

[23] M. A. Atianashie, "Artificial Intelligence (AI) Disease Detection in CCN-51 Cocoa Fruits through Convolutional Neural Networks: A Novel Approach for the Ghana Cocoa Board," Convergence Chronicles, vol. 5, no. 3. pp. 2955–7844, 2024. [Online]. Available: http://creativecommons.org/licenses/by/4.0/

[24] S. Ferraris, R. Meo, S. Pinardi, M. Salis, and G. Sartor, "Machine Learning as a Strategic Tool for Helping Cocoa Farmers in Côte D'Ivoire," Sensors, vol. 23, no. 17, 2023, doi: 10.3390/s23177632.

[25] P. Agbeli, K. Anokye, J. Kobina, and J. B. Hayfron-Acquah, "Application of Digital Image Processing Technology to Detect Diseases in Cocoa Plants." 2023. [Online]. Available: https://www.researchgate.net/publication/370873532

[26] D. Mamadou, K. J. Ayikpa, A. B. Ballo, and B. M. Kouassi, "Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms," International Journal of Advanced Computer Science and Applications, vol. 14, no. 9, 2023, doi: 10.14569/IJACSA.2023.0140937