LEAF-BASED CACAO DISEASES CLASSIFICATION USING IMAGE PROCESSING

Dominic O. Cagadas* and Ruth April A. Labajan
University of Science and Technology of Southern Philippines-Cagayan de Oro
Cagayan de Oro City, 9000, Philippines
*Correspondence: dominic.cagadas@ustp.edu.ph

ABSTRACT - *Cacao's diversified value as food and non-food ingredient increased its demand both in local and global markets; however, the presence of diseases affects its production. Identification of plant diseases is an important means of preventing losses in the yield and quantity of any agricultural crop, such as cacao. Manual means of identification and monitoring of diseases in cacao leaves is the common route, however, this requires a tremendous amount of work, expertise in plant diseases, and excessive processing time. Hence, an image processing technique in detecting cacao plant diseases through its leaves is presented in this paper. The cacao plant disease detection is conducted through the use of convolutional neural networks (CNN). The pre-trained CNN model, specifically the VGG19 model, is used as a feature extractor and classifier. Diseases under test are Swollen shoot, VSD (Vascular-Streak Dieback), and Witches Broom, in addition to healthy cacao leaves. Experimental results in this research show that the VGG19 model has achieved an accuracy of 88.75% for classifying cacao diseases through its leaves. The leaf-based classification via image processing of plant diseases in cacao could help farmers in early detection and intervention. However, an increased accuracy would be more viable if the number of images collected will be increased.*

Keywords: *Cacao Diseases, Convolutional Neural Network, VGG19 model, Swollen shoot, Vascular Streak Dieback, Witches Broom*

1. INTRODUCTION

Cocoa beans harvested from cacao plants are the main ingredient in the production of chocolate. Other than the chocolate industry, cocoa is gaining popularity in cosmetics and pharmaceutical industries because of its varied nutritional benefits; paving the way for broader market opportunities, and increasing further the supply and demand gap of cocoa beans in both local and global markets [1].

Despite the average increase of 2,743 ha per year from 2013 to 2020 in cacao production, supply is still considered low due to increased market demand, lack of knowledge of cacao growers, and most specifically, high mortality of planting material attributed to the presence of cacao diseases [1].

Cacao diseases contribute to more than a third of the potential crop, resulting in a decline in production as well as a reduction in the quality of beans in almost all the cacao-producing areas in the world [2]. Diseases in cacao plants reduces yield by 20% or 1.3 million tons of beans in 2012 [3]. The cocoa economy is under threat because of diseases that pose serious constraints to cocoa production [4].

Important cacao diseases are Vascular Streak Dieback (VSD) [5-8] which is characterized by yellowing of leaves [9]; Swollen-shoot Virus (CSSV) [10-13] characterized by red vein-banding in young leaves which may be followed by vein clearing or chlorosis along the veins [14]; and Witches' Broom [12, 15] characterized by formation of abnormal stems, flowers and pods [16].

Early detection of plant diseases for appropriate intervention is necessary as plant diseases place a huge factor in plant yield and productivity. Identification and detection of plant diseases are usually done by naked-eye observation by experts. This usually is not desirable as this entails cost and time. Thus, the importance of detecting plant diseases through automatic methods cannot be undermined as it reduces large work of monitoring especially in big farms; as well as detecting diseases at a very early stage when symptoms first appear on plant leaves [17]. The advent of technology led to the existence of methods that lead to faster and easier detection of plant diseases including remote sensing and image segmentation and soft computing [18]; and spectral vegetation indices [19]. In the Philippines, a study uses an image classification system interfaced with the drone to detect and monitor infestation in rice fields [20].

Lately, the use of deep learning in plant disease detection, among other agricultural operations [21] is slowly gaining a place. Deep learning (DL) algorithms have been shown to be extremely successful in real-world object detection, identification, and classification. A few state-of-the-art DL models have been used to perform plant disease detection through using DL architectures. The most commonly used approach for detecting plant diseases is convolutional neural networks (CNNs). It has shown excellent results in the area of image classification. For instance, a study [22] evaluates the results of various convolutional neural networks (CNN) in plant leaf diseases detection and classification such as in tea leaves [23]; cassava leaves [24]; tomato leaves [25]; groundnut diseases [26] and recently cacao diseases [27] in addition to the diversified application of image processing in agriculture such as in rubber seeds surface identification [28].

In today's world, the importance of accurate and timely disease detection, as well as early prevention, has never been greater. A variety of methods exist for detecting plant diseases as mentioned in the above studies. In cacao disease classification, CNN has been used [27]; however, only for VSD. In this paper, a plant disease recognition model based on leaf image classification, specifically for three important cacao leaf diseases (VSD, swollen shoot broom witch) using CNN VGG19 model was presented, as well as how it performs in terms of disease detection accuracy.

2. METHODOLOGY

This study focuses on cacao disease classification via image processing using cacao leaves. Leaves were chosen as the route of classification in this study since it is readily accessible, usable, and reusable and does not pose a threat to the plant when harvested. In addition, the bodily structure similarity between the underside of the leaves and the skin cells of the fruits justifies the use of leaves even more [29].

2.1 Image Acquisition and Resizing

Around 2000 images of cacao leaves were collected. Images of cacao leaves with diseases (VSD, CSSV, Witches' Broom) are collected on the internet while images of healthy leaves are collected in the field. Images were cropped and reshaped into desired image sizes. After this, images were classified according to their disease classes. The different classes for evaluation are presented in Table 1.

Table 1 Different Classes of Evaluation

Class	Description/Interpretation
Class 0	Healthy
Class 1	Swollen Shoot
Class 2	VSD
Class 3	Witches Broom

2.2 CNN VGG 19 Model

A pre-trained VGG19 CNN model was used for feature extraction and classification. The images are then fed to the CNN model system for training and validation.

In the CNN VGG19 model, data images were resized to 224x224 pixels and labeled into four groups/classes. Images are fed to the model after being preprocessed. The framework used a pre-trained VGG19 model. The two completely connected layers were replaced with a value of 128 for fine-tuning, and the Softmax classifier was replaced with four classes as needed in this analysis.

The CNN VGG19 model has a trainable parameter of 3,228,420 out of a total of 23,252,804 parameters. This model has almost 26 layers, each with a different number of convolution and max pooling layers, providing more training parameters. As a result, given a large number of training parameters, this model produces high accuracy.

2.3 Model Performance Measure

To measure the performance of the model, the following metrics were used with their corresponding equations adopted from [30], together with the application of the confusion matrix table:

- *accuracy*, which is the ratio between the total number of predictions produced and the number of right predictions made.

  Accuracy: $\frac{TN+TP}{FP+TN+TP+FN}$
- *precision,* which evaluates the accuracy with which a class is determined.

  Precision: $\frac{TP}{FP+TP}$
- *recall*, which pertains to the classifier's capacity to predict the class correctly.

  Recall: $\frac{TP}{FP+TP}$
- *F1 score*, which defines the harmonic sensitivity and accuracy value of the model.

  F1 Score: $2x\{\frac{Precision*Recall}{Precision+Recall}\}$
- *specificity rate*, which demonstrates the classifier's separation abilities.

  Specificity: $\frac{TN}{FP+TN}$

where:

TN- true negative;
TP- true positive;
FP – false positive;
FN – false negative

The confusion matrix shows the ratio between the predicted and the actual class.

3. RESULTS AND DISCUSSION

3.1 CNN VGG19 Model

Table 2 presents the classification report of the CNN VGG19 model with four classes. It is evident in the result that the average accuracy of the model is 88.75% across the four classes.

Table 2. Classification Report of CNN VGG19 Model

Class	Precision	Recall	F1 Score	Support
Class 0	0.84	0.93	0.89	92
Class 1	0.86	0.76	0.81	107
Class 2	0.96	.95	0.96	109
Class 3	0.88	0.91	0.89	92
Average	**0.89**	**0.89**	**0.89**	**400**

Precision is the percentage of predictions that are correctly classified, Recall is the percentage of positive cases over the whole number of the element, and F1 scores are the percentage of positive predictions that are correct. Table 2 shows that Class 2 (VSD) has the highest Recall with a percentage of 95% among the other results, this means that among 109 samples, 104 samples were identified to be true positive VSD disease while the rest were false negative (other classification). Class 2 (VSD) has the highest precision with 96% of predictions were correct, also 96% of positive predictions were correct (F1 score).

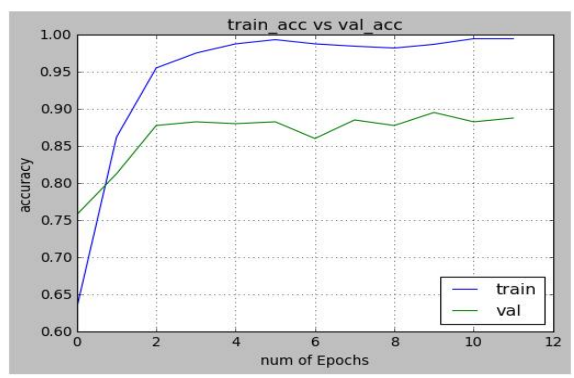

Figure 1. CNN Accuracy

Figure 2. CNN Loss

Figures 1 and 2 represent the graph model accuracy and loss of the CNN VGG19 model. It is evident from the graph that the system is stable, with accuracy consistent with the running number of epochs. Moreover, Figure 3 represents the confusion matrix of the CNN VGG19 model. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifiers") on a set of test data for which the true values are known.

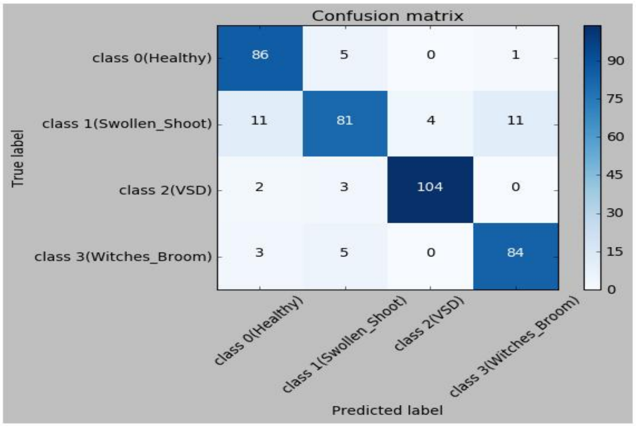

Figure 3. Confusion matrix of CNN VGG19 Model

In the confusion matrix table, it directly compares the actual target values with those predicted by the Vgg19 classifier. The highlighted values of each column are the values that are predicted to be true positive on its actual value. It is evident that the overall accuracy of the model is 88.75%.

It is shown that in Class 0 (healthy Leaves), 86 of the samples have been predicted out of 92, 5 samples have been predicted to be Class 1 (swollen-shoot), 0 on Class 2 (VSD), and 1 in Class 3 (witches' broom).

In the Class 1(swollen-shoot) column, 11 were classified as Class 0 (healthy), 81 samples in Class 1 (swollen-shoot), 4 samples in Class 2 (VSD), and 11 samples in Class 3 (witches' broom).

In the Class 2 (VSD) column, 2 samples were identified as Class 0 (healthy), 3 samples as Class 1 (swollen-shoot), 104 samples in Class 2 (VSD), and none were identified as Class 3 (witches' broom).

Last but not least column which is Class 3 (witches' broom), 3 samples were identified as Class 0 (healthy), 5 samples are Class 1 (swollen-shoot), none were identified as Class 2 (VSD), and 84 samples were identified as Class 3 (witches' broom) out of 92 samples.

3.2 Classification Prediction

Table 2 shows the different predictions of classifications. It shows different testing on the predictions from different inputs not included in the initial dataset.

The system will test each class whenever an input image is fed. The system will either predict or mis predict the correct classification. From the confusion matrix presented, in which the system resulted in 88.75% overall accuracy, it is evident from the class's prediction that the system will not accurately identify the system.

Table 2. Classification Predictions

	Image Under Test	Class0	Class1	Class2	Class3
Heal-thy		P	NP	NP	NP
		P	NP	NP	NP
		P	NP	NP	NP
Swo-llen shoot		NP	P	NP	NP
		NP	NP	NP	P
		NP	NP	P	NP
VSD		NP	NP	P	NP
		NP	P	NP	NP
		NP	NP	P	NP
Witches Broom		NP	NP	NP	P
		P	NP	NP	NP
		NP	NP	NP	P

P - Prediction Correct; NP - Prediction Incorrect

4. CONCLUSIONS AND RECOMMENDATIONS

This study has successfully implemented the leaf-based Cacao diseases recognition system using convolutional neural networks (CNN) at four classes namely: Witches Broom, Vascular Streak Dieback, Swollen Shoot, and Healthy Leaves. The VGG19 CNN Model was used to compare efficient and accurate results. As per the accuracy results of the VGG19, it is concluded that the optimal model for classifying cacao diseases under this study has an accuracy of 88.75%.

It is recommended in this study to collect more data images of leaves contaminated with the different cacao diseases to achieve higher accuracy. For commercial purposes, it is recommended to develop the system for mobile application as it will be the most accessible way of technology for the farmers especially those in remote places. Such technology will help them in the timely detection of cacao diseases contaminated in their farms would allow them to increase yield in farms by applying the right cure for the cacao disease infection, and to have a system that could show solutions on how to treat, or what fertilizers for the specific diseases.

5. REFERENCE

[1] "Philippine Cacao Roadmap (2021-2025)," Department of Agriculture-Bureau of Agricultural Research, 2022. Accessed: Jun. 11, 2023. [Online]. Available: https://shorturl.at/doAO7

[2] P.K. Hebbar, "Cacao diseases: A global perspective from an industry point of view," *Phytopathology*, **97**(12), 1658-1663 (2007).

[3] R. Ploetz, "The Impact of Diseases on Cacao Production: A Global Overview," *Cacao Diseases*, pp. 33–59, 2016, doi: https://doi.org/10.1007/978-3-319-24789-2_2.

[4] Dele Omoyele Adeniyi, "Diversity of Cacao Pathogens and Impact on Yield and Global Production," Nov. 2019, doi: https://doi.org/10.5772/intechopen.81993.

[5] G. J. Samuels *et al.*, "Vascular Streak Dieback of cacao in Southeast Asia and Melanesia: in planta detection of the pathogen and a new taxonomy," *Fungal Biology*, vol. 116, no. 1, pp. 11–23, Jan. 2012, doi: https://doi.org/10.1016/j.funbio.2011.07.009.

[6] D. G. Alvindia and F. L. M. Gallema, "Lasiodiplodia theobromae causes vascular streak dieback (VSD)–like symptoms of cacao in Davao Region, Philippines," *Australasian Plant Disease Notes*, vol. 12, no. 1, Sep. 2017, doi: https://doi.org/10.1007/s13314-017-0279-9.

[7]"Cocoa vascular streak dieback (223)," *apps.lucidcentral.org*. https://www.pestnet.org/fact_sheets/cocoa_vascular_streak_dieback_223.htm#:~:text=The%20first%20signs%20of%20the (accessed Jun. 11, 2023).

[8] D. Guest and P. J. Keane, "Cacao diseases: vascular-streak dieback," pp. 287–302, Aug. 2018, doi: https://doi.org/10.19103/as.2017.0021.18.

[9] T.C. Solpot, "Cacao *(Theobroma cacao*) diseases: threats to the Philippines cocoa industry," *Philippine Plant Disease Bulletin* 1(1), 2-4 (2020).

[10] E. Muller, "Cacao Swollen Shoot Virus (CSSV): History, Biology, and Genome," *Cacao Diseases*, pp. 337–358, 2016, doi: https://doi.org/10.1007/978-3-319-24789-2_10.

[11] H. Lot, E. Djiekpor, and M. Jacquemond, "Characterization of the genome of cacao swollen shoot virus," *Journal of General Virology*, vol. 72, no. 7, pp. 1735–1739, Jul. 1991, doi: https://doi.org/10.1099/0022-1317-72-7-1735.

[12] J.-P. Marelli *et al.*, "Chocolate Under Threat from Old and New Cacao Diseases," *Phytopathology®*, vol. 109, no. 8, pp. 1331–1343, Aug. 2019, doi: https://doi.org/10.1094/phyto-12-18-0477-rvw.

[13] G. A. Ameyaw, "Management of the Cacao Swollen Shoot Virus (CSSV) Menace in Ghana: The Past, Present and the Future," *Plant Diseases - Current Threats and Management Trends*, Apr. 2020, doi: https://doi.org/10.5772/intechopen.87009.

[14] A. F. Posnette, "VIRUS DISEASES OF CACAO IN WEST AFRICA: I. CACAO VIRUSES 1A, 1B, 1C AND 1D," vol. 34, no. 3, pp. 388–402, Sep. 1947, doi: https://doi.org/10.1111/j.1744-7348.1947.tb06372.x.

[15] G.W. Griffith, "Witches' brooms and frosty pods: threats to world cacao production," The Biologist, 51(2), 71-75 (2004).

[16] H. C. Evans, "Pleomorphism in Crinipellis perniciosa, causal agent of witches' broom disease of cocoa," *Transactions of the British Mycological Society*, vol. 74, no. 3, pp. 515–523, Jun. 1980, doi: https://doi.org/10.1016/s0007-1536(80)80051-9.

[17] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, Mar. 2017, doi: https://doi.org/10.1016/j.inpa.2016.10.005.

[18] J. L. Hatfield, A. A. Gitelson, J. S. Schepers, and C. L. Walthall, "Application of Spectral Remote Sensing for Agronomic Decisions," *Agronomy Journal*, vol. 100, p. S-117-S-131, May 2008, doi: https://doi.org/10.2134/agronj2006.0370c.

[19] A.-K. . Mahlein *et al.*, "Development of spectral indices for detecting and identifying plant diseases," *Remote Sensing of Environment*, vol. 128, pp. 21–30, Jan. 2013, doi: https://doi.org/10.1016/j.rse.2012.09.019.

[20] G.Z. Gamboa Jr., A.S. Morite, R.B. Bacarro, R.A. Plando, V.J. Ylaya, and E.J. Serna, "Ricefield Health Monitoring System Using a Drone with AI Interface," *Sci. Int. (Lahore),* 35(3), 181-184, 2023.

[21] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018, doi: https://doi.org/10.1016/j.compag.2018.02.016.

[22] B. Tugrul, E. Elfatimi, and R. Eryigit, "Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review," *Agriculture*, vol. 12, no. 8, p. 1192, Aug. 2022, doi: https://doi.org/10.3390/agriculture12081192.

[23] R. Singh, N. Sharma, and R. Gupta, "Proposed CNN Model for Tea Leaf Disease Classification," May 2023, doi: https://doi.org/10.1109/icaaic56838.2023.10140680.

[24] Prashant Giridhar Shambharkar and S. Sharma, "Plant Disease Detection And Prevention Using Deep Learning," Mar. 2023, doi: https://doi.org/10.1109/icaccs57279.2023.10112733.

[25] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," *Procedia Computer Science*, vol. 167, pp. 293–301, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.225.

[26] M. P. Vaishnnave, K. Suganya Devi, and P. Ganeshkumar, "Automatic method for classification of groundnut diseases using deep convolutional neural network," *Soft Computing*, Apr. 2020, doi: https://doi.org/10.1007/s00500-020-04946-0.

[27] Annisa Fitri Maghfiroh Harvyanti, Rifki Ilham Baihaki, None Dafik, Zainur Rasyid Ridlo, and Ika Hesti Agustin, "Application of Convolutional Neural Network for Identifying Cocoa Leaf Disease," pp. 283–304, Jan.

2023, doi: https://doi.org/10.2991/978-94-6463-174-6_21.

[28] S.M. Johari, S.K. Bejo, and W.W. Ishak, "Identification of dorsal and vertical surface of rubber seeds using image processing approach". *Sci Int (Lahore)*, *29*(2), pp.121-125, 2017.

[29] S. Nyassé, C. Cilas, C. Herail, and G. Blaha, "Leaf inoculation as an early screening test for cocoa (Theobroma cacao L.) resistance to Phytophthora black pod disease," *Crop Protection*, vol. 14, no. 8, pp. 657–663, Dec. 1995, doi: https://doi.org/10.1016/0261-2194(95)00054-2.

[30] Taspinar, Y. S., Koklu, M., and Altin, M., "Fire Detection in Images Using Framework Based on Image Processing, Motion Detection, and Convolutional Neural Network," *International Journal of Intelligent Systems and Applications in Engineering*, **9**(4), 171-177(2021).