

Machine Learning Techniques in the Detection of Cocoa (*Theobroma cacao* L.) Diseases

Ciro Rodriguez¹, Oswaldo Alfaro², Pervis Paredes³, Doris Esenarro⁴, Francisco Hilario⁵

¹Universidad Nacional Mayor de San Marcos UNMSM, Lima Peru

^{1,2,3,4,5}Universidad Nacional Federico Villarreal UNFV, Lima Peru

ABSTRACT

The purpose of the research is to apply machine learning techniques to identify the cocoa tree's diseases (*Theobroma cacao* L.) and avoid the loss of crop harvests because farmers lack immediate tools to detect diseases on time. The methodology considers the use of machine learning with techniques for image processing and analysis such as HoG (Histograms of Oriented Gradient), LBP (Local Binary Pattern), and the SVM (Support Vector Machine) algorithm, for the classification to determine if the plant cocoa is being affected or not by disease. The results obtained show that SVM, Random Forest, and ANN's application with the characteristic vectors extracted with the HOG and LBP extraction algorithms predict the cocoa plant state; therefore, it is advisable to increase the dataset so that the results are more accurate.

Keywords: Machine Learning, Cocoa, HOG ((Histograms of Oriented Gradient), LBP (Local Binary Pattern), SVM (Support Vector Machine), ANN (Artificial Neural Networks)

Introduction

Exports of cocoa in Peru grew 11%, the most demanded being its presentation in beans. The countries that buy the most Peruvian product are the Netherlands and the United States. Followed by Indonesia, Germany, Malaysia, Belgium, Italy, Mexico, Spain, and others. [1]

Peru has become a major global food supplier with nearly US \$ 5 billion in annual shipments. In the last two decades, agro-export companies have outlined growth strategies either through the diversification of crops and markets or production areas. This effort leads to the ranking in exports of asparagus, mangoes, grapes, cocoa, and other products that mostly consume fresh foods.

Companies are also experimenting with the uses of Peruvian cocoa, a crop that has begun to be revalued by the world market in the last five years due to its high quality and aroma. Along these lines, the Amaz Foods firm is marketing a craft beer Cacao Brown Ale made with fine aroma cacao nibs from the San Martín and Amazonas regions. The alcoholic beverage is marketed in the local market, but they do not rule out other markets with this novel proposal. [2]

According to the Ministry of Agriculture and Irrigation (Minagri), more than 100,000 families, mainly family farming, are engaged in cocoa cultivation in 16 of the 24 regions of the country. Cocoa is grown in 16 of the 24 departments, and in 2018 the national production was 135,000 tons over 160,000 hectares.

The main cocoa production areas are located in San Martín, Junín, Ucayali, Cusco, Huánuco, Amazonas, and Ayacucho, representing 93% of the total national production.

According to the International Cocoa Organization (ICCO), 75% of Peruvian exports correspond to fine and aroma cocoa, being a differentiating attribute compared to other types of cocoa. In the same way, 90% of cocoa production and its preparations are destined for export, mainly to the US and European markets.

Peru remains the ninth world producer of cocoa beans and the second world producer of organic cocoa. [3]

The most common diseases that affect cocoa production can become a severe problem, compromising a large part of the harvest and the plant's life. Due to the magnitude of losses caused by diseases, the farmer has substituted other crops for cocoa.

It is a significant need for producers to control diseases that affect cocoa crops, the disease that most affects cocoa production is the black pod which acts the harvest of this in a relatively high percentage.

The causes of these diseases are poor management of the plantations, the producers' abandonment, and the lack of technical knowledge applicable to the field, on the other hand, to the bad weather conditions in certain regions. Another factor that causes these diseases' proliferation is that there are no norms and laws regulating producers who have infected plantations without taking sanitary control measures, causing infections to farms nearby. [4]

State of the Art

For this research, other related research works using image machine learning techniques were taken as references. The thesis "Assessment of Internal and External Quality of blueberries using images" was based on the fact that, during the storage and transport of blueberries, some defects in the fruit used to cause rejections during their commercialization, for this reason, it was believed that carrying out some methodological studies Non-destructive for external and internal evaluation during postharvest would be valuable. There were two main axes: the external evaluation of the quality of the postharvest blueberries with pattern recognition techniques using color images, and the evaluation of the internal quality of the blueberries using hyperspectral images, to build predictive models. . Statistical pattern recognition was implemented to detect three of the most common defects in blueberries: fungus, dehydration shrinkage, and blueberries with compressed impacts and two main orientations (pedicel and calyx).

Images of the blueberries classified in the orientations and defects mentioned above were acquired to extract chromatic and geometric characteristics and then make a selection for which a sequential forward search algorithm was used; the selected features were used for the training of different classifiers to determine which of them reported a better classification, some classifiers used were: linear analysis discriminant, vector support machine, and probabilistic neural network, these being the ones that also reported better performance after training and pattern recognition. For evaluating the internal quality, predictive models were built from the hyperspectral images in two detection modes: reflectance and transmittance; their spectral information was used. This technique was based on image processing and spectroscopic analysis, resulting in multiple images with spectral information suitable for building models. Therefore, in this thesis, it was demonstrated that using images using computer vision for external evaluation and hyperspectral images to measure internal quality is feasible to improve the product's quality and commercialization. [5].

The other research that was taken as a reference was the thesis "Evaluation of Classifiers for Automatic Disease Detection in Citrus leaves using machine vision" this research was proposed to evaluate an adequate control for the diseases presented in the citrus industry because this industry is important in the agricultural economy of Florida, it is one of the main producers of jobs for the residents of this state. Every year large amounts of chemicals are used as fungicides to control different diseases that affect citrus crops, generating enormous costs for producers. In this thesis, a study was implemented that investigated the use of artificial vision and image processing techniques to classify diseased citrus leaves. Four different classes were taken into account: leaves with greasy spots, melanosis leaves, regular leaves, and leaves with scabies. Leaf image data was collected using a JAI MV90, 3 CCD color camera with 28-90 zoom lens.[6][7][8]

In addition, algorithms based on image processing techniques were used for feature extraction. The feature extraction process used the color co-occurrence methodology (CCM method). This method consists of taking into account both the color and the texture of the images you want to work on to obtain unique characteristics. In this study, the classifiers were: statistical classifier using the Mahalanobis minimum distance method, neural network based on the backpropagation algorithm, and neural network using radial basis functions. He studied determined that such sorting methods are suitable for sorting citrus leaves [9]. This research was a feasibility analysis to demonstrate whether the investigated techniques can detect diseases in citrus plants. The results obtained showed whether they could be used for this application; however, the analysis of the images for this research was carried out in a laboratory so that in outdoor conditions, the taking of the images would face other challenges such as natural lighting. And the structure of the tree would be an obstacle to overcome so that the research carried out is applicable for the classification of leaves of citrus plants based on the image.[10][11][12].

Methodology

It is proposed to implement a disease recognition model in the leaves of cocoa plants. This proposal seeks to solve the problems that arise in cocoa crops with presence since with this model it would be easy and fast, without having to take it to a laboratory for analysis and recognition, in this way, the detection of the type of disease that the plant presents, just by obtaining an instantaneous capture or a short video of the plant and with it, the system would recognize the type of disease with one that it has from the first capture. Thus it could be treated in time and would not generate losses to the cocoa farmers.[13][14][15]

It is hoped to obtain a disease recognition system that will detect whether the plant is sick or not. The work methodology used in this research is the one observed in Figure 1. The first phase to be carried out is the input of the photographs' database, photos of the cocoa leaves with different diseases were collected.[16][17][18]

The images were taken with different devices to have captured with different resolutions and images of different qualities.

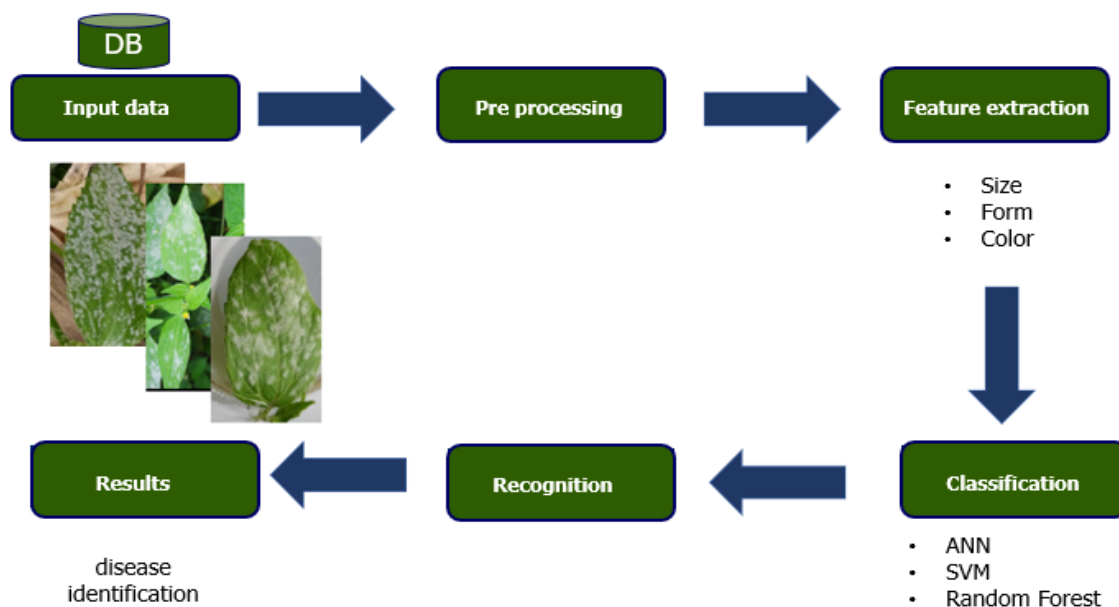


Figure 1: Work methodology

Some examples of the diseases that can be identified in cocoa leaves:

- Alternariasp: It is a fungus that reproduces rapidly on dry leaves.
- Whitefly: They are located in the back of the plants' leaves, and they eat them.
- Fruit borer (*Heliothis virescens*): The larvae pierce fruits contaminated by their feces and pathogens.
- Thrips (*Thripstabaci*): It feeds on the floral parts and interferes with the fruit's pollination and binding.
- Sooty mold: It appears initially as a thin layer of black color.

Development

To achieve the objective of characterizing the most common types of diseases that affect cocoa cultivation, it was necessary

- Identify the main characteristics of the diseases that affect cocoa plants
- Identify the effects that cause diseases in the cocoa plant.
- Report the results of the system for the application of preventive measures.

The classification of diseases was carried out considering:

The diseases are generally caused by fungi and pathogens of rapid reproduction transmitted through contact between plants and man, animals, and the wind's action. Table 1 shows the types of diseases.

Table 1.Types of diseases

Disease	Fungi / pathogens
Moniliasis.	Moniliophthoraroreri
The black ear.	Phytophthora complex
Machete disease.	Ceratocystisfimbriata
Anthrachnose.	Colletotrichumgloeosporoides Penza
The bubas.	Cushions affect

Moniliasis.

Caused by the fungus *Moniliophthoraroreri*, which attacks the pods or cocoa fruits at any age, causing rotting of the beans. The severity of the attack varies according to the area, time of year, and climate.

The black ear.

Caused by fungi of the *Phytophthora* complex, it is responsible for more losses in crops; the fungus can attack seedlings and different parts of the cocoa tree, such as flower cushions, suckers, shoots, leaves, branches, trunk, and roots, the main damage has suffered the ears. It is considered the most critical disease in 80% of cocoa producing countries.

Machete disease.

Caused by the *Ceratocystisfimbriata* fungus that destroys entire trees, the fungus infects cacao through lesions on the main trunks and branches and can kill a tree quickly. If trees killed by this disease are not well controlled, it can cause a loss of dead trees of up to 10% over several years.

Anthrachnose.

Caused by the fungus *Colletotrichumgloeosporoides* Penza, also called by the common name Anthrachnose where it has been distributed worldwide. In cocoa, the fungus attacks the stem, leaves, suckers, and fruits. Damage to pods is not economically important, although injury to stalks is.

The bubas.

Bulging and abnormal growth of flower cushions affects cocoa plants, especially vegetative shoots, flower cushions, and young fruits; in short, it attacks meristematic tissues (young) in active growth. The scientific name knows it of (*Crinipellis perniciososa*).

Most common cocoa plant disease symptoms

Moniliasis.

The disease's external symptoms are that a chocolate-colored spot is formed where all the grains and tissues are already affected. Table 2 shows the

Table 2. Cocoa symptoms with Moniliasis disease.

Disease	Plant disease symptoms
Moniliasis	• Light or greasy spots on the cob
	• Lumps, Gibas or chichotes
	• Premature Maturity
	• Bright brown or chocolate-colored stain.

Table 3: Disease Severity Scale Developed by Horsfall and Heuberger [19]

Category	Severity
0	Apparently infected
1	10 – 25% leaf Area infected
2	26 – 50% leaf Area infected
3	51 – 75% leaf Area infected.
4	> 75% leaf Area infected

- Light or greasy spots on the cob.

The greasy spots are hard to see, but if you do a good ear check, they can be seen on the green ears; the spots are yellow. In the red fruits, the dots are orange to check that the ear is infected; peel the shell with the machete where the light spot is observed, and if you notice little brown dots, the fungus has already entered the fruit. On many occasions, there is already rotting of the seeds.

- Lumps, Gibas or chichotes

The bumps or humps appear in the first two months of the fruit's age, the gherkin stage of the ear, and appear due to the fungus' entry into the fruit. When the ears with hump or chichote are left on the ground, the disease's development stops, and no spores are produced.

- Premature Maturity

Early maturity is a misleading symptom since you can believe that it has ripened when looking at an ear, but you find that the seeds are rotten when you open it. Fruits with this symptom are heavier than healthy ripe ears.

- Bright brown or chocolate-colored stain.

The chocolate brown spots grow irregularly until they completely cover the ear. As the stain advances, the fruit becomes heavier and begins to dry out. The ears left on the ground in the chocolate stain state produce spores or seeds that can infect other fruits for up to 20 days.

- White and Creamy Powder (spores or seeds)

The last symptom in a fruit attacked by *Monilia* is the chocolate stain of a whitish powder that later turns creamy. This appears 6 or 10 days after the chocolate stains and is very dangerous

since this dust is the spores or seeds of the fungus, which can infect other healthy fruits or plantations. Ears left on the ground with white powder can infect other healthy ears and plantations for 30 days.

Techniques

HoG (Histograms of Oriented Gradient) - Robert K. McConnell: Used to detect computer vision and image processing objects. Counts the orientation gradient occurrences in localized portions of an image - detection window or region of interest (ROI). [7][16][17]

The implementation of the HoG algorithm are as follows:

- A detection window within the frame.
- A color normalization is made.
- The gradients of the subimage are obtained to calculate the histogram of candidates on uniform cells.
- Cells are grouped into blocks with some overlap.
- Cell blocks are normalized independently.
- The feature vector is made up of a set of block descriptors.

LBP (Local Binary Pattern)

- It is a method used in the extraction of texture characteristics for classification reasons.
- It has the characteristic that it is invariant to lighting changes in the gray levels.
- It is ideal for applications that require rapid feature extraction and texture classification.

After extracting all the images' characteristics, the separations will be made into two classes: test and test, because for the training of the model, we will work with the test class and verify the model we will use the test class.

The algorithm to be used is the SVM (Support Vector Machine); this algorithm is a classification method in which each data is graphed as a point in the space of n dimensions (where n is the number of variables you have) With the value of each variable being the value of a particular coordinate, this algorithm will determine if the cocoa plant is being affected by a disease, or not.

Dataset

The dataset used for this project has been taken based on a collection of still increasing data.

The data used for this project were collected in a farm plantation in northern Peru; the images were taken over six months and stored in a database adding the disease's characteristics as shown in Figure 2 and Figure 2, with farmer's support's experience and knowledge.

Image properties

File type: JPG file.

Dimensions: 256 * 256.

Width: 256 pixels.

Height: 256 pixels.



Figure 2. Disease cocoa image



Figure 3. Healthy cocoa image

The algorithms using Machine Learning techniques were developed with the Python language and the open source OpenCV library for artificial vision.[18][19][20]

```
data_dir = '/FarmNorthen'
train_dir = data_dir + '/train'
valid_dir = data_dir + '/val'
nThreads = 4
batch_size = 32
use_gpu = torch.cuda.is_available()
defimg_segmentation(rgb_img,hsv_img):
    light = np.array([25,0,20])
    lumps = np.array([100,255,255])
    healthy_mask = cv2.inRange(hsv_img, light, lumps)
    result = cv2.bitwise_and(rgb_img,rgb_img, mask=healthy_mask)
    premature = np.array([10,0,10])
    bright = np.array([30,255,255])
    disease_mask = cv2.inRange(hsv_img, premature, bright)
    disease_result = cv2.bitwise_and(rgb_img, rgb_img,mask=disease_mask)
    final_mask = healthy_mask + disease_mask
    final_result = cv2.bitwise_and(rgb_img, rgb_img, mask=final_mask)
    returnfinal_result
```

Results

The results obtained with the SVM, Random Forest, and Neural Networks models and tested with the characteristic vectors extracted with the HOG and LBP extraction algorithms are shown in Table 4.[21-25]

Table 4: Tabla de resultados esperados

	SVM	RandomForest	ANN
HOG	0.70	0.50	0.65
LBP	0.75	0.60	0.85

La Tabla 5, muestra los resultados de identificación de los síntomas de la enfermedad del cacao con la Miniliasis y se presenta el porcentaje de la deteccion.[26-30]

Table 5. Cocoa symptoms with Moniliasis disease.

Disease	Plant disease symptoms	% detection
Moniliasis	• Light or greasy spots on the cob	75%
	• Lumps, Gibas or chichotes	69.5%
	• Premature Maturity	72.3%
	• Bright brown or chocolate-colored stain.	73%

Table 6 shows the precision percentages obtained after applying the automatic learning techniques. The results obtained from processing the image samples according to the infection rates and the degree of the disease show an approximation of 75%.

Table 6. Disease classification accuracy

Sample image	Machine Learning Technique		
	% infected	Disease grade	Accuracy disease
1	13.323	0.75324	75%
2	4.756	0.75456	75%
3	10.456	0.64323	74%
4	7.324	0.55034	75%
5	3.435	0.65034	75%
6	2.768	0.45034	74%
7	7.543	0.55034	75%
8	5.432	0.65034	73%
9	10.434	0.45034	74%
10	3.435	0.55034	72%

It is expected that with the model trained with the neuronal network algorithm and using the characteristic vectors extracted with LBP, the state of the cocoa plant can be predicted; it is recommended that you work with large amounts of data so that the result is more accurate.

Conclusión

According to the revised investigations, it is concluded that the model has a correct prediction; that is, it can be accurate when defining the class to which a new image inserted in the model belongs (whether it is the cocoa plant with a disease or not). It is recommended that the disease images database be larger to process the recognition results to be reliable. You can work with other feature extraction algorithms such as sift and surf, which would be used with deep learning to choose the algorithm that gives the best results.

References

1. M. Ortiz, «La agroindustria busca conquistar los mercados del mundo con innovación,» 2017. [En línea]. Available: <https://www.redagricola.com/pe/la-agroindustria-busca-conquistar-los-mercados-del-mundo-innovacion/>. [Último acceso: 13 Octubre 2020].
2. ADEX, «Feria Expoalimentaria Abrir sus Puertas,» Peru Exposrta, vol. Edicion Especial, n° 416, p. 64, Julio 2019.

3. AgenciaPeruana de Noticias, «En Perú más de 100,000 familias se dedican al cultivo de cacao,» ANDINA, 1 Octubre 2019.
4. Y. Jaimes y F. Aranzazu, «MANEJO DE LAS ENFERMEDADES DEL CACAO (Theobroma cacao L) EN COLOMBIA, CON ÉNFASIS EN,» UNIÓN TEMPORAL DE CACAO - COLOMBIA DOS, p. 90, 2010.
5. A. N. Rathod , B. A. Tanawala, V. H. Shah. 2014. “Leaf disease detection using image processing and neural network”, IJAERD vol. 1, no. 6, June, pp. 1-10.
6. S. Jayaraman, S. Esakkirajan , and T. Veerakumar. 2009. Digital Image Processing. U.S.A.: McGraw Hill Co.,.
7. H. Rizk, «Automated early plant disease detection and grading system:,» AUC Knowledge Fountain., Cairo, 2017.
8. J. Fierrez-Aguilar, N. Alonso-Hermira, G. Moreno-Marquez, and J. Ortega-Garcia, “An off-line signature verification system based on fusion of local and global information,” in International workshop on biometric authentication, pp. 295–306, Springer, 2004.
9. P. G. S. Silva, C. A. M. L. Junior, E. B. Lima, B. L. D. Bezerra, and C. Zanchettin, “Speeding-up the handwritten signature segmentation process through an optimized fully convolutional neural network,” in 15th International Conference on Document Analysis and Recognition (ICDAR 2019), vol. 1, IEEE, 2019.
10. X. Zhang, X. Zhou, M. Lin, and J. Sun, “Shufflenet: An extremely efficient convolutional neural network for mobile devices,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856, 2018.
11. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention, pp. 234– 241, Springer, 2015.
12. S. S. Patki, G. S. Sable. 2016. “Cotton Leaf Disease Detection & Classification using Multi SVM,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 10, pp. 165-168.
13. A. Lundgren, D. Castro, E. B. Lima, and B. L. D. Bezerra, “Octshufflenet: A compact octave based neural network for end-to-end multilingual text detection and recognition,” in 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), vol. 4, pp. 37–42, IEEE, 2019.
14. V. A. Gulhane, A. A. Gurjar. 2011. “Detection of disease on cotton Leaves and Its Possible Diagnosis”, International Journal of Image Processing (IJIP), vol. 5 : no 5 : pp. 590-598.
15. S. S. Panchal, R. Sonar. 2016. “Pomegranate Leaf Disease Detection Using Support Vector Machine,” International Journal of Engineering and Computer Science, vol. 5, no 6, pp. 16815-16818.
16. Y. Vargas; R. H. Puerta; F. Palomino Vera; C. Rodriguez; B. Pandey. Low planting densities for early maturation of *Mauritia flexuosa* for sustainable plantations in Alto Huallaga, Peru. Dec 2020 in World Journal of Engineering. doi: 10.1108/WJE-09-2020-0416
17. V. M. Tiwari, T. Gupta. 2017. “Plant Leaf Disease Analysis using Image Processing Technique with Modified SVM-CS Classifier,” International Journal of Engineering and Management Technology, vol. 5, no. 1, pp. 11-17.
18. MyoMyo Han. 2017. “Analysis on Detection Improvement of Textile Fabric Defects using Gray Level Co-occurrence Matrix for automatic system,” M.E. thesis, Dept. Electronic Eng., Yangon Technological University., Yangon, Myanmar.

19. Bustamante, J.C., Rodriguez, C., Esenarro, D. (2019). Real time facial expression recognition system based on deep learning. *International Journal of Recent Technology and Engineering*, 2019, 8(2 Special Issue 11), pp. 4047-4051. doi: 10.35940/ijrte.B1591.0982S1119
20. D. Gadkari. 2004. "Image Quality Analysis Using GLCM," M.E. thesis, University of Central Florida, Orlando, Florida.
21. V. K. S. L. Melo and B. L. D. Bezerra, "A fully convolutional network for signature segmentation from document images," in 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 540–545, IEEE, 2018.
22. J. C. Sharman, "Power and discourse in policy diffusion: Antimoney laundering in developing states," *International Studies Quarterly*, vol. 52, no. 3, pp. 635–656, 2008.
23. C. A. B. MELLO, W. P. SANTOS, and A. L. I. OLIVEIRA, *Digital Document Analysis and Processing*. New York: Nova Science Publishers Inc, 2012.
24. P. Forczmanski, A. Smoliński, A. Nowosielski, and K. Małecki, "Segmentation of scanned documents using deep-learning approach," in *International Conference on Computer Recognition Systems*, pp. 141–152, Springer, 2019.
25. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
26. L. Hamers et al., "Similarity measures in scientometric research: The jaccard index versus salton's cosine formula.," *Information Processing and Management*, vol. 25, no. 3, pp. 315–18, 1989.
27. Z. Cai, X. He, J. Sun, and N. Vasconcelos, "Deep learning with low precision by half-wave gaussian quantization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5918–5926, 2017.
28. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.
29. KoKoZaw, Zin Ma MaMyo, and WahWahHlaing. 2018. "Multiclass Support Vector Machine Based Detection and Classification of Leaf Diseases", 11th National Conference on Science and Engineering, YTU, Yangon, Myanmar
30. A. Miyatra, S. Solanki. 2014. "Disease and Nutrient Deficiency Detection in Cotton Plant," *International Journal of Engineering Development and Research*, vol. 2, no 2, pp. 2801-2804.