ORIGINAL ARTICLE

Improved YOLOv5m model based on Swin Transformer, *K-means++*, and Efficient Intersection over Union (EIoU) loss function for cocoa tree (*Theobroma cacao*) disease detection

Benedicta Nana Esi Nyarko*, Wu Bin Wu, Zhou Jinzhi Zhou, Mwanaharusi Mohd Juma Mohd

School of Information Engineering, Southwest University of Science and Technology, Qinglong Road, Fucheng District, Mianyang C, Mianyang, China

Abstract

The cocoa tree is prone to diverse diseases such as stem borer, stem canker, swollen shot, and root rot disease which impedes high yield. Early disease detection is a critical component of diverse management processes that are implemented throughout the life cycle of cocoa plants. Consequently, several studies on the application of detection techniques to recognize diseases have been proposed by several researchers. This study proposes the YOLOv5m network for cocoa tree disease detection. The development of cocoa disease detection systems will aid farmers in early identification prompt response, and efficient management of related cocoa tree diseases which will ultimately increase yield and sustainability. To improve the performance of the YOLOv5m network, a Swin Transformer (Swin-T) was added to the backbone network to improve cocoa tree disease detection accuracy. By obtaining global information, the *K-means++* algorithm was added to modify the choice of initial clustering locations, and Efficient Intersection over Union Loss (EIoU) was used as a bounding box regression loss function to speed up the bounding box regression rate, resulting in higher precision of the YOLOv5m network. The experimental assessment outcome of this study showed that the proposed method YOLOv5m (Swin-T, *K-means++*, EIoU) achieved 96% precision, mAP of 92%, and recall of 94%. Compared to the original YOLOv5m, precision improved by 5%, mAP improved by 6%, and recall by 5%. Comparing the proposed method to the conventional YOLOv5m, the latter showed improved performance and better accuracy with a high detection speed and compactness. This improvement offers a useful and effective method for detecting diseases related to cocoa trees.

Keywords: cocoa tree disease, EIoU, K-*means++*, plant disease detection, Swin Transformer

Introduction

A crucial area of research in the field of deep learning is the detection of pests and plant diseases (Shoaib *et al.* 2023). It is a technology that gathers images of plants using computer vision equipment to determine the existence of diseases in plants. Currently, deep-learning-based tools for identifying plant diseases and pests have largely replaced traditional naked-eye methods of identification in the agricultural sector (Bhatti *et al.* 2023). Plant diseases and pest detection in real complex natural environments is difficult due to factors like low contrast, wide variations in the scale of the lesion area and different types, and a lot of noise in the lesion image. Deep learning models have seen success in recent years in a variety of computer vision applications (Gong and Zhang 2023). Numerous deep learning-based plant diseases and pest detection techniques are used in smart agricultural practice, and businesses have created a range of deep learning-based plant disease detection applications (Nandini 2023).

Theobroma cacao, also known as the cacao tree or cocoa tree, is a small, evergreen tree in the Malvaceae family (Debnath *et al.* 2023). It is primarily found in

the Amazon basin and the Guyana Plateau in neotropical rainforests. It is a typical cauliflorous plant that is not strictly monoecious, is cross-pollinated and grows to a height of 6–12 m (20–39 ft). The fruit, also referred to as cocoa pod, is a kind of indehiscent berry that weighs an average of 500 g but can vary depending on the clone. It has between 30 and 60 sweet mucilage--coated seeds inside (Snell 2023).

For millions of farmers in Africa, Asia, Central, and South America, it provides a substantial and reliable base of agricultural income. Cacao is distinguished by a significant genetic diversity, and its seeds are the primary raw material for the multi-billion-dollar chocolate industry (Appiah 2023). Over the years, cocoa has been a significant economic crop and a major source of income for many cocoa farmers (Okali 2023), but its production has drastically decreased due to low yields brought on by crop diseases and farmers' failure to maintain healthy crops (Appiah 2023). Like most other tropical perennial plants, the cocoa tree also heavily depends on the availability of genetic resources from wild sources for genetic improvement, particularly for disease resistance standards (Argout *et al.* 2023).

The cocoa tree is prone to diverse diseases that impede high production such as stem borer, stem canker, swollen shot, and root rot disease (Ambele *et al.* 2023). Stem canker is a pathogenic disease that affects cocoa caused by *Phytophthora palmivora* and *P. megakarya* (Bose *et al.* 2023). Cocoa stem canker is considered to be the third most destructive disease that affects Cocoa swollen shoot virus disease and black pod disease (Nyadanu *et al.* 2012). Several parts of the cocoa tree exhibit stem canker infections. Grayishbrown, water-soaked lesions on the bark with a wide dark brown to black edge are the outward signs of the condition (Appiah 2023). These pathogens cause cankers by infecting bark, flower cushions, and chupons. Such lesions leak a reddish-brown liquid that eventually dries and turns into a rusty deposit (Adeniyi and Asogwa 2023).

The cocoa swollen shoot virus (CSSV), which belongs to the *Genus Badna* virus, is what causes CSSVD. Only seen in West Africa, CSSVD has been documented in Togo, Sierra Leone, Ghana, Nigeria, and Côte d'Ivoire (Muller 2016). CSSVD is considered the most financially significant cocoa infection and causes 15–50% yield loss depending on the severity of the strain (George 2019). Cocoa stem borer (cocoa moth) is caused by *Eulophonotus myrmeleon* (Kingsley-Umana *et al.* 2022). The larvae of the moth feed on the stem, thereby resulting in a loss in production yield. Recently, the implementation of integrated management practices has led to the adoption of computer-based systems for the early detection of cocoa pod borer infestation (Bahadur and Dutta 2023).

The root rot fungus affects a variety of hosts, including cocoa plantations and native forest trees. Additionally, it affects breadfruit, and it has been linked to the atoll-country-specific Pingelap disease (Puig 2023; Angira *et al.* 2024). It was first recorded in coconut in Papua New Guinea. Starting at the tips of the branches, the fungus' attack on the roots causes the leaves to turn yellow and wilt. All the leaves disappear very quickly. On the trunk, the fungus develops a crust that can reach a height of 1 meter. It is dark brown to black with a white margin, and frequently contains clear liquid drops (Bissiri *et al.* 2023).

Among the various management processes, early disease detection is crucial. To this end, several studies have reported the use of detection methods to identify swollen shoot diseases. Ramos-Sobrinho *et al.* (2021) investigated molecularly swollen shoot disease in cocoa produced in Togo by amplification with four PCR primer pairs. In a similar study, Ameyaw *et al.* (2018) investigated the variable detection of CSSVD polymerase chain reaction amplification. Recently, computerbased algorithms have been applied in the detection of SSDs (Yan 2022). For instance, Coulibaly *et al.* (2020) studied the detection of swollen shoot diseases using convolutional neural networks. Hacinas *et al.* (2022) investigated the detection of the cocoa moth by employing a deep learning algorithm based on edge computing. Similar studies have been conducted on different crops. For instance, in a study by Bhandarkar *et al.* (2019), the detection of borers in coffee was done using statistical models and deep learning. This study explored related cocoa tree diseases and proposed the YOLOv5m model for disease detection. The model was enhanced by incorporating the Swin Transformer (Swin-T) into the model's backbone to obtain multiscale characteristics. The *K-means++* algorithm complemented the model by altering the initial clustering location selection which increases the model's precision. The Efficient Intersection over Union Loss (EIoU) was used as a bounding box regression loss function to accelerate the bounding box regression rate. The integration of these models into the YOLOv5m network significantly increased the efficiency of the network.

The development of disease detection systems in cocoa trees is essential for different purposes. This research can aid farmers in taking immediate action to stop the spread of diseases to healthy plants by detecting diseases early on. This can safeguard the general well-being of cocoa plantations and drastically lower yield losses. Systems for detecting diseases produce data that can be used to analyze and comprehend disease trends, environmental factors that affect the prevalence of diseases, and the efficiency of disease management techniques. For long-term prevention and control of disease, this data-driven approach facilitates

informed decision-making. Detecting cocoa tree diseases is primarily done to minimize negative effects on the environment and to enhance the welfare of communities that grow cocoa, all the while ensuring the sustainability, profitability, and quality of cocoa farming operations.

Materials and Methods

This section delineates the data and general workflow of techniques adopted for the application of deep learning methods in cocoa tree disease detection.

Dataset analysis and preprocessing

The cocoa tree disease detection dataset contains a total of 1,000 images. The datasets were annotated and preprocessed using the Roboflow software. They were then downloaded in YOLOv5 format resulting in a total of 2173 with four (4) classes namely: swollen shoot, stem borer, brown root rot, and healthy stem. The labeled image names are grouped into training, test, and validation sets with a 7 : 2 : 1 ratio as shown in Figure 1.

Experimental setup and design

The hardware environment configuration for this experiment was carried out on a Lenovo laptop with an Intel Core i7 2.50 GHz processor and NVidia GeForce GTX 860M GPU, and the software environment was CUDA 11.0, CUDNN version 8, windows 11 operating systems, python 3.8.0, and visual studio. This study proposed the YOLOv5m network for cocoa tree disease detection. The performance of the YOLOv5m network was improved by adding a Swin Transformer to the backbone network to acquire multi-scale features. By obtaining global information, the *K-means++* algorithm was added to modify the choice of initial clustering locations, and Efficient Intersection over Union

Loss (EIoU) was used as a bounding box regression loss function to speed up the bounding box regression rate, resulting in higher precision of the YOLOv5m network. Ablation experiments were performed to ascertain the improvement of the network performance.

The YOLOv5 model

A highly effective network structure in the single-stage object detection network is the YOLO series network (Wang *et al.* 2023). The model backbone acquires features of various sizes, assembles these attributes through the neck, and then generates these maps P3, P4, and P5, to detect small, medium, and large objects (Diwan *et al.* 2023). Four modules make up its network architecture: input, backbone, neck, and head. To make this algorithm faster and more accurate, it incorporates some new ideas for improvement based on YOLOv4. Images are sent to the backbone for feature extraction after being processed through the input layer. YOLOv5's structure is logically divided into four main sections: input, neck, backbone, and prediction.

The input dataset is preprocessed in the input section using methods like image scaling, responsive anchor box processes, and data augmentation. Both Focus and CSP structures are incorporated into the YOLOv5 backbone network. The Focus module's purpose is to segment the input image to increase network speed and increase the local receptive field range (Bacea and Oniga 2023).

The CSP1_X and CSP2_X are two distinct designs that are part of the CSP structure. These architectures improve the network's capacity for learning, ensuring accuracy while cutting down on operations. The CSP allows the model to capture a richer set of features by dividing the original input into two different sections that are each subjected to convolution operations. The model can capture a richer set of features thanks to CSP, which divides the original input into two parallel branches and subjects each to convolution operations. This backbone relies heavily on the CBL module,
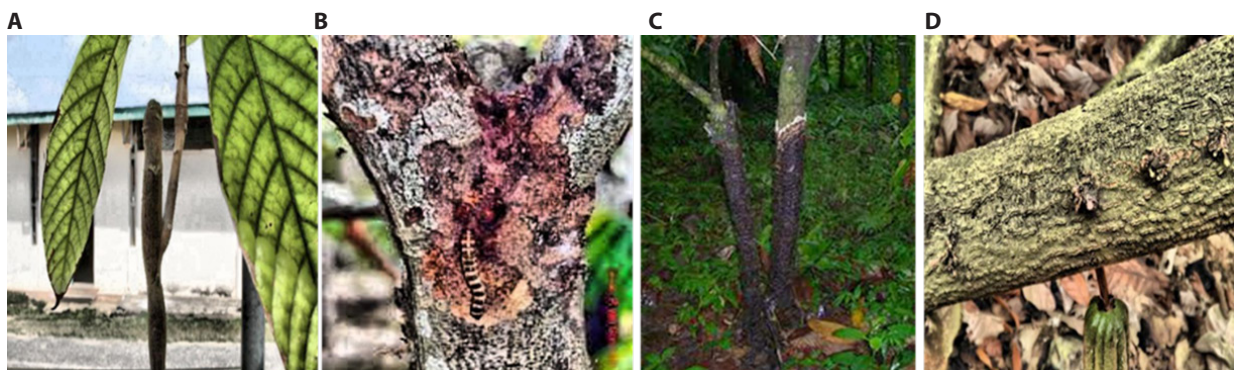

Fig. 1. Sample cocoa tree dataset: A – swollen shoot, B – stem borer, C – root rot, D – healthy stem

which stands for Conv + BN + LeakyRelu. To guarantee effective feature extraction and network proliferation, this module integrates a leaky rectified linear unit (LeakyRelu) activation function, a convolutional layer, and a batch normalization layer. To harmonize features across distinct layers, the Neck section employs a unique feature pyramid network (FPN) + path aggregation network (PAN) structure. PAN consolidates high and low features via up-and-down sampling operations, while FPN layers transmit semantic features in a top-down manner. Finally, two crucial components of object detection are covered in the Prediction segment: non-maximum suppression and bounding box loss. Following the delivery of the three feature maps to the prediction head, each pixel in the feature map is subjected to bounding box regression and confidence calculation using the preset a priori box to determine the object category, category confidence, and box coordinates. multidimensional arrays with data on width and height. Upon executing the non-maximum suppression (NMS) procedure and establishing the correlating thresholds to eliminate superfluous data from the array, the ultimate detection information can be generated.

With CSPDaknet53 as the backbone. The primary architecture consists of stacking multiple C3 and CBS (Conv + BatchNorm + SiLU) modules, followed by the connection of an SPPF module. The SPPF module improves the backbone's ability to express features, and the CBS module helps the C3 module with feature extraction. SPPF uses the max pooling function before max pooling to avoid repeating SPP operations in SPP-Net which turns to accelerate the module's speed. The parameters of the YOLOv5 backbone are illustrated in Table 1.

Table 1. Parameters of YOLOv5 backbone

Module	Parameters	Arguments
CBS	3520	[3,32,6,2,2]
CBS	18,560	[32,64,3,2]
C3	18,816	[64,64,1]
CBS	73,984	[64,128,3,2]
C3	115,712	[128,128,2]
CBS	295,424	[128,256,3,2]
C3	625,152	[256,256,3]
CBS	1,180,672	[256,512,3,2]
C3	1,182,720	[512,512,1]
SPPF	656,896	[512,512,5]

YOLOv5m architecture enhancement

YOLOv5 comes in five different versions: YOLOv5x, YOLOv5l, YOLOv5m, YOLOv5s, and YOLOv5n. The smallest model in the YOLOv5 series is the YOLOv5s. "S" represents "small", mobile or edge devices. They have limited processing power and are the ideal platforms for this model's performance. Although YOLOv5s has the fastest detection speed, it is not very accurate. The YOLOv5 series also includes a medium-sized model. "M" represents "medium". YOLOv5m is appropriate for devices with specific computing capabilities and offers a good balance between speed and accuracy. The YOLOv5x is the largest model in the YOLOv5 family. "X" represents extra-large. Although YOLOv5x has the slowest detection speed, it performs best in terms of accuracy and is ideal for tasks needing a very high level of accuracy and for devices with strong computing powers (GPUs, for example). The
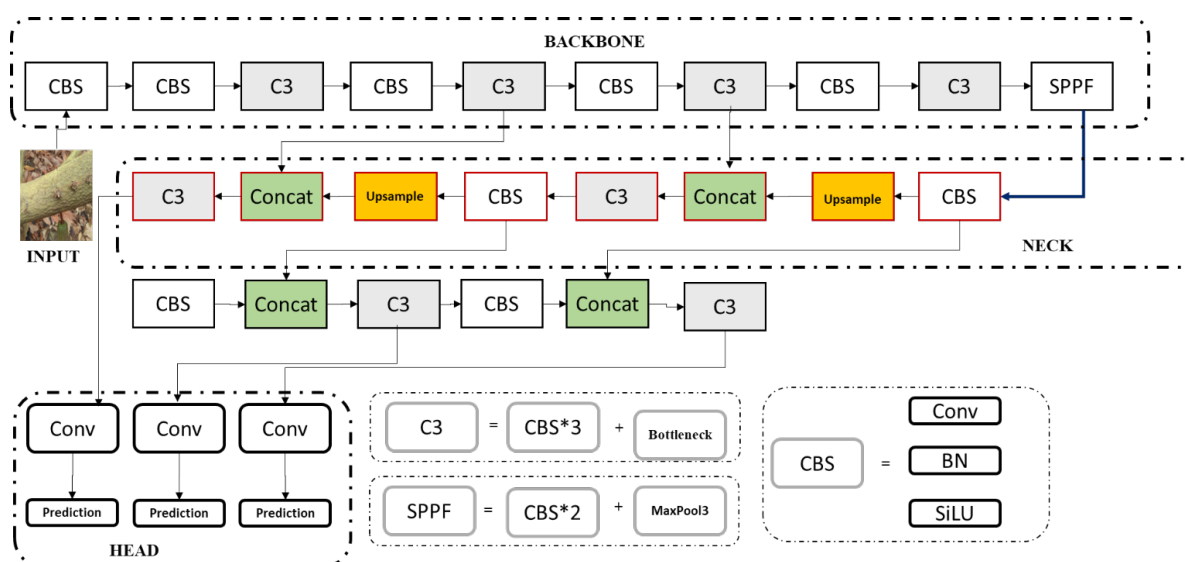
Fig. 2. Structure of the YOLOv5 architecture

YOLOv5n is an improved version of the YOLOv5 series designed for nanodevices like NVIDIA JETSON Nano. This version of the YOLOv5 series maintains speed while offering edge-detection accuracy. In summary, several YOLOv5 versions are tailored to meet varying demands for processing power and real-time performance. The primary factors to consider while selecting a suitable model are the device's processing capacity and the task's accuracy needs. After successfully training YOLOv5x, YOLOv5l, YOLOv5m, YOLOv5s, and YOLOv5n on the cocoa tree datasets, the YOLOv5m was maintained for improvements due to its outstanding performance over other variants.

This research enhanced the YOLOv5m network structure specifically for cocoa tree disease detection. A Swin Transformer was integrated into the backbone network. Furthermore, the Efficient intersection over union (EIoU) Loss function was added to boost performance. Finally, appropriate initial anchor box positions were determined by applying the *K-means++* clustering algorithm to modify the choice of initial clustering locations.

K-means++ for anchor box optimization

Although *K-means* is used by the original YOLOv5 to cluster anchor box positions, *K-means* has built-in drawbacks. Initially, *K-means* is a heuristic approach, so it does not guarantee convergence to a global optimum. Second, the choice of the starting center has a direct impact on the clustering result. *K-means* can easily result in local fusion or necessitate more iterations because it randomized selections of sample points as cluster centers. To ensure the best anchor boxes for cocoa tree-related disease detection, the *K-means++* algorithm was proposed to improve the initial clustering center selection. *K-means++* handles the target point sampling like a probability-based task, in contrast to *K-means*, which ensures the increase in speed of the algorithm convergence. Due to its stochastic setup phase, which lowers the possibility of poor convergence and guarantees a more reliable and effective clustering process, *K-means++* is superior to *K-means*. Using a weighted likelihood probabilistic model, the *K-means++* algorithm selects a new data point at random to serve as the new center. When each point is chosen, $P(x)$, is computed in Equation (1) as:

$$P_{(x)} = \frac{D_{(x)}^2}{\sum_{x \in X} D_{(x)}^2} . \quad (1)$$

Loss function optimization

The GIoU loss function is utilized in standard YOLOv5 configurations to quantify bounding box prediction incongruities. GIoU, however, has limitations that result in less-than-ideal training convergence, particularly when bounding boxes have different aspect ratios or are misaligned. When this happens, GIoU falls back to a simple intersection over union (IoU) metric, which is insufficient to accurately represent the actual amount of intersection between the bounding boxes. The Efficient Intersection over Union (EIoU) loss function was proposed to offer a steady gradient and is more adept at managing aspect ratio fluctuations and misalignments. The EIoU loss function breaks down the aspect ratios of both the predicted and actual bounding boxes, in contrast to GIoU, which has issues with bounding boxes that are misaligned or have drastically different aspect ratios. This allows for individual computations of width and height and produces a more accurate and subtle depiction of the overlap between bounding boxes. The EIoU loss function is mathematically expressed as:

$$EIoU = (1 - iou) + \alpha \cdot d_c + \beta \cdot (d_w + d_h). \quad (2)$$

Equation 2 uses the intersection-to-union area ratio (IOU) to quantify the correlation between the ground truth and predicted bounding boxes.

Coefficient α measures the Euclidean distance between the centerline of the accurately predicted and ground truth bounding boxes d_c which serve as a modulatory weighting factor for the center distance loss. Similarly, β serves as a weighting coefficient for the dimensional loss terms and, $d_w + d_h$ which represent the differences in height and width, respectively, between the minimally circumscribed bounding box and the predicted bounding box. EIOU improves regression accuracy and speeds up convergence for anticipated bounding boxes. This gets around GIOU's built-in limitations and guarantees stable learning across a range of object sizes and shapes.

The Swin Transformer

To output multi-scale feature information, patches are continuously merged at each stage, starting with a smaller patch size and using feature maps with varying down sampling rates at different stages. Swin Transformer is more appropriate for a range of computer vision tasks such as plant disease detection due to its hierarchical network structure for acquiring multi-scale features. These features enable the network to perform intensive object detection tasks more effectively. A pretrained model of the Swin T was used in the YOLOv5 backbone of the cocoa tree dataset. The initial learning rate Adam was set to 0.01 with 0.0005 weight decay. The learning rate was then decreased to 0.00001 at the 20th epoch and 0.000001 at the 50th epoch, respectively. During the training epochs, the YOLOv5m + Swin-T achieves 89% precision

at the 50th epoch compared to YOLOv5m's 0.8383% precision at the 50th epoch. This shows 0.6% significant improvements of the Swin-T when added to the YOLOv5m backbone architecture.

The first Swin Transformer block, which has a two-layer MLP with GELU non-linearity and two two-layer norm layers, was based on W-MSA, as seen in Figure 3. Each W-MSA module and MLP module was followed by a residual connection and a Layer Norm layer, respectively. The second Swin Transformer block was identical to the first one, with the exception that a SW-MSA block took the place of the W-MSA block.
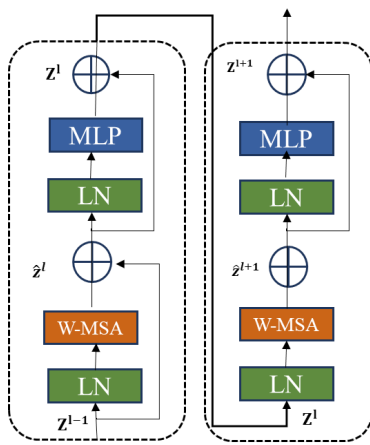

Fig. 3. Swin Transformer block

Shifted Window-based Multi-head Self-Attention (SW-MSA), which forms a new image by shifting the initial one both vertically and horizontally, was introduced by Swin Transformer. The Swin Transformer block is mathematically expressed in Equation 3 as:

$$\begin{cases} \hat{z}^l = W - MSA\left(LN(z^{l-1})\right) + z^{l-1} \\ z^l = MLP\left(LN(\hat{z}^l)\right) + \hat{z}^l \\ \hat{z}^{l+1} = SW - MSA\left(LN(z^l)\right) + z^l \\ z^{l+1} = MLP\left(LN(\hat{z}^{l+1})\right) + \hat{z}^{l+1} \end{cases} \quad (3)$$

Results and Discussion

Experimental evaluation criteria

mAP50 was the experiment's evaluation criterion in this paper. The average of all classification detection results' AP50 values is known as mAP@50. When the IOU threshold is set at 0.5, the AP50 value denotes the closed region of the precision and recall curve.

Equations 4–7 present the computational analysis for precision, recall, and mAP@50:

$$Recall = \frac{TP}{TP + TN} \quad (4)$$

$$AP@50 = \int_0^1 P_r d_{r,\ IOU} \quad (5)$$

$$AP@50 = \int_0^1 P_r d_{r,\ IOU\ \geq 0.5} \quad (6)$$

$$mAP = \frac{1}{C}\sum_{i=1}^{n} P(k)\Delta R(k) \quad (7)$$

True Positives (TP) denote the number of targets that the model detected; False Positives (FP) indicate the number of targets that the model incorrectly detected; and False Negatives (FN) indicate the number of targets that the model missed during detection. C denotes the number of disease target categories, n denotes the number of IoU thresholds, k is the IoU of $P(k)$ is precision, and $R(k)$ is Recall. mAP@0.5 refers to the average AP of all classes when IoU is set to 0.5.

Results

Three sets of comparative experiments were carried out in this study The first experiments compared the YOLOv5m model with other YOLOv5 variants such as YOLOv5n, YOLOv5s, YOLOv5l, and YOLOv5x. Both the task's accuracy requirements and the device's processing capabilities are important considerations for choosing an appropriate model. Based on our system's hardware configurations as stated in YOLOv5m architecture enhancement, the above variants of the YOLOv5 model were analyzed by training each variant with the cocoa tree disease dataset. Table 2 detailed the optimization techniques for each variant and the results obtained after training each variant at 50 epochs. Model name defines the name of the YOLOv5 variant. Size defines the image resolution, and params define

Table 2. Experimental results of YOLOv5 models

Model	Epochs	Precision	Recall	mAP_0.5	mAP_0.5:0.95
YOLOv5m	50	0.91	0.89	0.86	0.86
YOLOv5n	50	0.88	0.89	0.84	0.88
YOLOv5s	50	0.90	0.90	0.66	0.72
YOLOv5l	50	0.74	0.73	0.71	0.73
YOLOv5x	50	0.70	0.72	0.69	0.68

the parameters corresponding to each model variant. Precision refers to the percentage of all detection results that are accurately detected. Recall is used to show the degree to which a positive input is used to make positive prediction. Epochs refer to a complete pass through the dataset. mAPval@0.5:0.95 represent the mean average precision at different IoU threshold from 0.5 to 0.95. mAPval@0.5 calculates the average precision of all images in each category, params define the parameters corresponding to each model variant.

The second experiment compared the performance of each proposed method (Swin Transformer, *K-means++*, and EIoU) with the YOLOv5m model. Applying the Swin Transformer to the YOLOv5m network global information was obtained. The *K-means++* modifies the choices of the initial clustering location, and the EIOU speeds up the bounding box regressions rate. The above modification made to the YOLOv5m model resulted in a significant improvement in the model's performance in terms of precision and recall. Sample results are presented in Table 3 and Figure 4, respectively. To ensure the validity of the test results, each set of comparative experiments applied the same training set, test set, training strategy, and evaluation criteria.

The trained results showed that YOLOv5m achieved the best prediction results of 91% against other models that achieved 85%, 90%, 74%, and 70%, respectively. The YOLOv5n and YOLOv5s took a short time to reach 50 epochs with low performance. YOLOv5l and YOLOv5x also took multiple hours to reach 50 epochs with low precision accuracy. The best

precision was achieved by YOLOv5m in an average time. The above analysis determined the proposed method to increase YOLOv5m Precision and detection accuracy. Hence the study proposed a Swin Transformer to the YOLOv5m network to obtain the global information. The *K-means++* was also added to modify the choices of the initial clustering location, and the EIOU sped up the bounding box regressions rate. An analysis of the experimental results showed that there was significant improvement in the proposed model in detecting cocoa tree disease. The results in Table 2 influenced the choice of the model we chose to improve. From Table 3 it can be seen that the YOLOv5m with Swin Transformer, *K-means++*, and EIoU achieved a significant improvement with an average increase of 5%.

The performance of the upgraded model was compared to YOLOv5m. To further evaluate the efficacy of our approach (the improved model), a comparison with comparable recent crop disease detection research is essential. Using computer vision, the suggested model YOLOv5m (EIOU + *K-means++* + Swin-T) was compared with the approach proposed by Coulibaly *et al.* (2020), Rodriguez *et al.* (2021), and Kumi *et al.* (2022). Coulibaly *et al.* (2020) proposed a way to extract features effectively to enhance better diagnosis of plant disease. Rodriguez *et al.* (2021) proposed machine-learning techniques for cocoa tree disease detection. Kumi *et al.* (2022) designed a mobile application with machine learning techniques integrated for cocoa pod-related disease detection. The results obtained by the proposed methods of these authors are tabulated in Table 5. The results illustrated in the table below show

Table 3. Experimental results of improved YOLOv5m model

Method	Epochs	Precision	Recall	mAP_0.5	mAP_0.5:0.95
Improved YOLOv5m + (Swin-T, *K-means++*, and EIoU)	50	0.96	0.94	0.92	0.93
YOLOv5m+ *K-means++*,	50	0.92	0.91	0.90	0.89
YOLOv5m + EIoU	50	0.94	0.91	0.92	0.92
YOLOv5m + Swin-T	50	0.95	0.90	0.90	0.91

Table 4. Comparative analysis results of the improved models at each epoch

| Model	20 Epochs			30 Epochs			40 Epochs			50 Epochs		
	precision	recall	mAP_0.5	precision	recall	mAP_0.5	precision	recall	mAP_0.5	precision	recall	mAP_0.5
YOLOv5m + (Swin-T, *K-means++*, and EIoU)	0.58	0.56	0.55	0.64	0.63	0.65	0.88	0.89	0.87	0.96	0.94	0.92
YOLOv5m + Swin-T	0.54	0.53	0.51	0.75	0.74	0.74	0.79	0.77	0.78	0.89	0.84	0.85
YOLOv5m + *K-means++*	0.54	0.52	0.50	0.68	0.66	0.66	0.79	0.78	0.79	0.86	0.84	0.84
YOLOv5m + EIoU	0.52	0.50	0.56	0.55	0.56	0.54	0.78	0.80	0.77	0.87	0.86	0.81
YOLOv5m	0.56	0.56	0.54	0.66	0.64	0.65	0.89	0.86	0.82	0.83	0.89	0.88
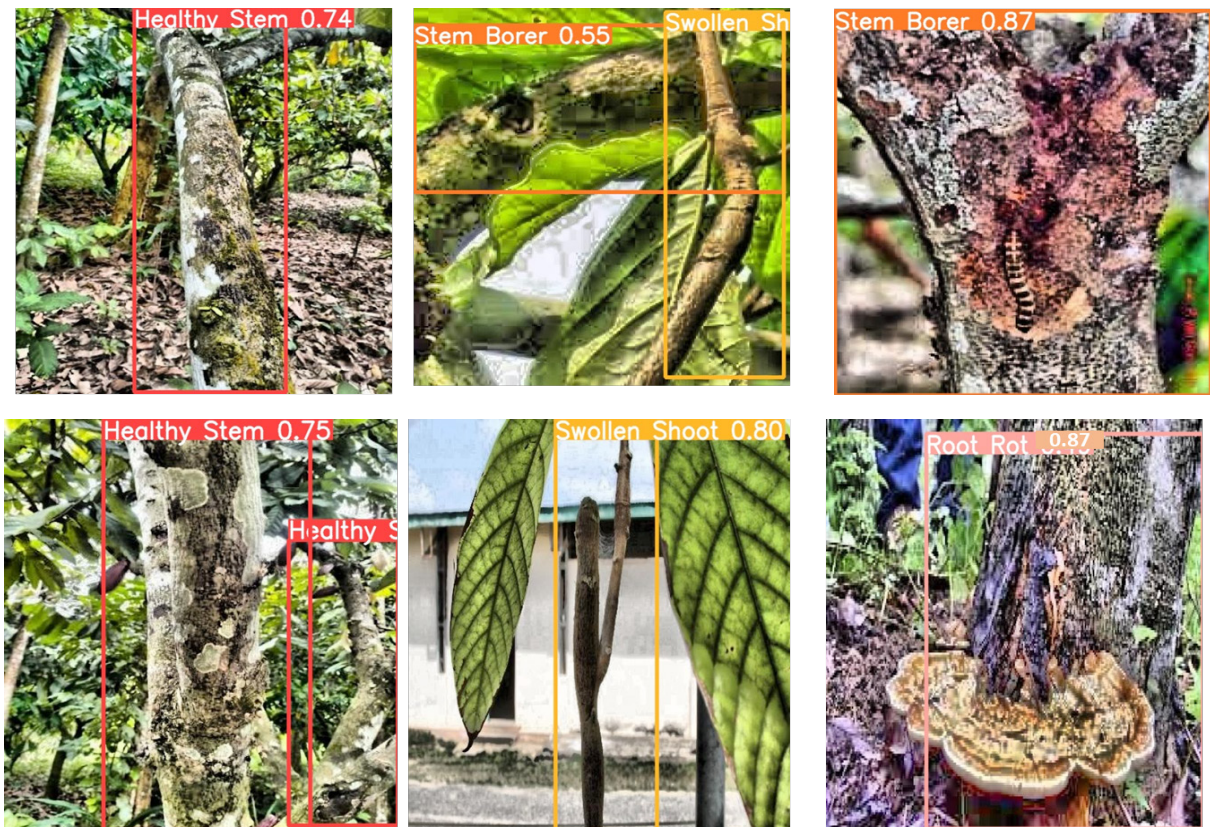
Fig. 4. Sample test results

Table 5. Comparative analysis of experimental results with other studies

Authors	Approach	Description	Dataset Size	Number of classes	Precision
Our method	YOLOv5m (EIOU + *K-means*++, + Swin-T)	cocoa tree disease detection	1,000	4	95.16
Coulibaly *et al.* (2020)	(DWT + PCA + GLCM + CNN)	cocoa tree disease detection	10,000	2	74%
Rodriguez *et al.* (2021)	proposed CNN model	cocoa pod disease detection	N/A	3	75%
Kumi *et al.* (2022)	mobile application with ML techniques	cocoa pod disease detection	N/A	N/A	80%

that our method achieved higher precision than Coulibaly *et al.* (2020) Rodriguez *et al.* (2021), and Kumi *et al.* (2022).

Discussion

The rapid technological development of the agricultural industry has been aided by the industry's adoption of a wide range of new technological solutions in recent years. Although agriculture is expected to become the industry standard, the physical risks and hazards that exist in this sector are significant factors that could prevent these innovations from being widely adopted and implemented. The assessment of many issues in agriculture has been made possible in recent years by the integration of artificial intelligence techniques. Current research has focused on developing deep-learning techniques for image acquisition at various levels to address plant detection issues. Numerous diseases, such as root rot, swollen shot, stem borer, and stem canker, impede the high production of cocoa trees. Early disease detection is crucial for a variety of management processes. As a result, diverse research on the application of detection methods for disease identification has been suggested.

Kumi *et al* (2022), proposed a deep learning-based smartphone application for cocoa pod disease detection. Their proposed approach incorporates a mobile

application with machine learning techniques. It allows cocoa farmers to snap a photo of the cocoa pod and submit it for diagnosis, which is done via a backend cloud service. Convolutional Neural Networks (CNN) are the foundation for automatic disease diagnosis and detection in image analysis and classification.

Given that the average recall for a few of the models is suitable for practical use, they concluded that the experimental results are encouraging and should be taken into consideration, in particular, SSD Resnet50 V1 FPN and SSD MobileNet V2 results. They also intend to publish an expanded version of this work in the future, along with a detailed analysis of each ML framework's performance on the dataset.

Rodriguez *et al.* (2021) used machine learning techniques to determine the diseases affecting cocoa trees (*Theobroma cacao* L.). The methodology uses machine learning with image processing and analysis techniques like the SVM (Support Vector Machine) algorithm, LBP (Local Binary Pattern), and HoG (Histograms of Oriented Gradient) for classification to determine whether or not the cocoa tree is affected by disease. The results obtained indicate that the cocoa plant state can be predicted by applying SVM, Random Forest, and ANN with characteristic vectors extracted using the HOG and LBP extraction algorithms. As a result, expanding the dataset is recommended to improve the accuracy of the results.

Conclusions

This study proposed YOLOV5m for cocoa tree disease detection. By integrating Swin Transformer, *K-means++*, and EIOU strategies, the YOLOv5m demonstrated enhanced performance with the various methods compared to the original YOLOv5m while maintaining high detection speed and compactness. This study contributes a practical and efficient solution for cocoa tree-related disease detection.

However, there remain some restrictions in our current research. First, the total number of related disease data obtained is still limited when compared to the number of cocoa tree-related diseases. In future research, more cocoa tree-related diseases will be incorporated as a detection target to expand the models' capabilities. In addition, while the improved YOLOv5m achieves a balance between precision and efficiency, using a more advanced network could potentially further boost detection performance. All things considered, this work provides a strong baseline model for cocoa tree disease detection, and future research will concentrate on increasing category diversity, optimizing accuracy, and enhancing generalizability.

Despite the rapid advancements in object detection, plant disease detection remains a challenge Implementation may be hampered by variables like transition costs, model generalization, dataset restrictions, and efficiency limitations. We will focus our future efforts on addressing these issues to enable wider implementation of object detection algorithms in plant disease detection. The current efforts made by diverse research in plant disease detection will gradually close the gap between cutting-edge methods and real-world applications.

References

Adeniyi D.O., Asogwa E.U. 2023. Chapter 14 – Complexes and diversity of pathogens and insect pests of cocoa tree. Forest Microbiology. Tree Diseases and Pests 3: 285–311. DOI: https://doi.org/10.1016/B978-0-443-18694-3.00002-X

Ambele C.F., Bisseleua H.D., Djuideu C.T., Akutse K.S. 2023. Managing insect services and disservices in cocoa agroforestry systems. Agroforestry Systems: 1–20. DOI: https://doi.org/10.1007/s10457-023-00839-x

Ameyaw G.A., Chingandu N., Domfeh O., Dzahini-Obiatey H.K., Gutiérrez O.A., Brown J.K. Variable detection of cacao swollen shoots disease-associated badnaviruses by PCR amplification. p. 13–17. In: "Proceedings of International Symposium on Cocoa Research (ISCR)". Lima, Peru. DOI: https://doi/full/10.5555/20203127100

Angira A., Sharma T., Choudhary N., Rishi N., Verma H.N., Awasthi L.P. 2024. Viral diseases of field and Horticultural Crops. Academic Press: 15–21. DOI: https://doi.org/10.1016/B978-0-323-90899-3.00015-X

Appiah A.A. 2023. Evaluation of progress in cocoa crop protection and management. IntechOpen: 216. DOI: https://doi.org/10.5772/intechopen.112642

Argout X., Droc G., Fouet O., Rouard M., Labadie K., Rhoné B., Rey Loor G., Lanaud C. 2023. Pangenomic exploration of *Theobroma cacao*: new insights into gene content diversity and selection during domestication. bioRxiv: 2023.2011.2003.565324. DOI: https://doi.org/10.1101/2023.11.03.565324

Bacea D.-S., Oniga F. 2023. Single-stage architecture for improved accuracy real-time object detection on mobile devices. Image and Vision Computing 130: 104613. DOI: https://doi.org/10.1016/j.imavis.2022.104613

Bahadur A., Dutta P. 2023. Diseases of commercial crops and their integrated management. CRC Press: 38–52. DOI: https://doi.org/10.4324/9781032627908

Bhatti U. A., Bazai S. U., Hussain S., Fakhar S., Ku C. S., Marjan S., Yee P. L., Jing L. 2023. Deep learning-based trees disease recognition and classification using hyperspectral data. Computers, Materials & Continua 77 (1): 681–697. DOI: https://doi.org/10.32604/cmc.2023.037958

Bissiri Y.H.-T., Ollo S., Maurice K.L.M., Senan S., Kolo Y. 2023. Socio-entomological characterisation of cocoa orchards in the haut-sassandra region (Côte d'Ivoire). Journal of Entomology and Zoology Studies 11 (3): 42–48. DOI: https://doi.org/10.22271/j.ento.2023.v11.i3a.9203

Bose T., Spies C.F., Hammerbacher A., Coutinho T.A. 2023. Phytophthora: an underestimated threat to agriculture, forestry, and natural ecosystems in sub-Saharan Africa. Mycological Progress 22 (11): 78. DOI: https://doi.org/10.1007/s11557-023-01926-0

Coulibaly M., Kouassi K., Kolo S., Asseu O. 2020. Detection of "swollen shoot" disease in Ivorian cocoa trees via convo-

lutional neural networks. Engineering 12: 166–176. DOI: https://doi.org/10.4236/eng.2020.123014

Debnath A.J., Dutta P., Bahadur A. 2023. Diseases of commercial crops and their integrated management. CRC Press: 38–52 DOI: https://doi.org/10.4324/9781032627908

Diwan T., Anirudh G., Tembhurne J.V. 2023. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimedia Tools and Applications 82 (6): 9243–9275. DOI: https://doi.org/10.1007/s11042-022-13644-y

George A.A. 2019. Management of the cacao swollen shoot virus (CSSV) menace in Ghana: The past, present and the future. Plant Diseases. T.-P. Snježana. IntechOpen. Rijeka 1–13 DOI: https://doi.org/10.5772/intechopen.87009

Gong X., Zhang S. 2023. An analysis of plant diseases identification based on deep learning methods. The Plant Pathology Journal 39 (4): 319. DOI: https://doi.org/10.5423%2FPPJ.OA.02.2023.0034

Hacinas E.A., Querol L., Acero L.A., Arcelo M., Amalin D., Rustia D.J. 2022. Automated cocoa pod borer detection using an edge computing-based deep learning algorithm. In: Proceedings of "ASABE Annual International Meeting". Houston, Texas, USA. DOI: https://doi.org/10.13031/aim.202200238

Kingsley-Umana E., Asogwa E., Mokwunye I. 2022. Outbreak, distribution and damage characteristics of cocoa stem borer, eulophonotus myrmeleon felder 1874 (Lepidoptera: Cossidae) in major cocoa producing states in Nigeria. Advances in Entomology 10: 175–185. DOI: https://doi.org/10.4236/ae.2022.102013

Kumi S., Kelly D., Woodstuff J., Lomotey R.K., Orji R., Deters R. 2022. Cocoa companion: Deep learning-based smartphone application for cocoa disease detection. Procedia Computer Science 203: 87–94. DOI: https://doi.org/10.1016/j.procs.2022.07.013

Muller E. 2016. Cacao swollen shoot virus (CSSV): history, biology, and genome. p. 337–358. In: "Cacao Diseases" (Bailey B., Meinhardt L., eds.). Springer, Cham DOI: https://doi.org/10.1007/978-3-319-24789-2_10

Nandini M.L.N. 2023. Detection of plant diseases and pests using deep learning models: a recent research. Chelonian Research Foundation 18 (2): 474–501. DOI: https://doi.org/10.18011/2023.10(2).474-501

Nyadanu D., Akromah R., Adomako B., Kwoseh C., Lower S.T., Dzahini-Obiatey H., Akrofi A.Y., Assuah M.K. 2012. Inheritance and general combining ability studies of detached pod, leaf disc and natural field resistance to *Phytophthora palmivora* and *Phytophthora megakarya* in cacao (*Theobroma cacao* L.). Euphytica 188 (2): 253–264. DOI: https://doi.org/10.1007/s10681-012-0717-x

Okali C. 2023. Female and male in West Africa. Routledge: 169–178 DOI: https://doi.org/10.4324/9781003402749

Puig A.S. 2023. Fungal pathogens of cacao in Puerto Rico. Plants 12 (22): 3855. DOI: https://doi.org/10.3390/plants12223855

Ramos-Sobrinho R., Kouakou K., Bi A. B., Keith C. V., Diby L., Kouame C., Aka R. A., Marelli J.-P., Brown J. K. 2021. Molecular detection of cacao swollen shoot badnavirus species by amplification with four PCR primer pairs, and evidence that cacao swollen shoot Togo B virus-like isolates are highly prevalent in Côte d'Ivoire. European Journal of Plant Pathology 159 (4): 941–947. DOI: https://doi.org/10.1007/s10658-021-02203-0

Rodriguez C., Alfaro O., Paredes P., Esenarro D., Hilario F. 2021. Machine learning techniques in the detection of cocoa (*Theobroma cacao* L.) diseases. Annals of the Romanian Society for Cell Biology 25 (3): 7732–7741.

Shoaib M., Shah B., Ei-Sappagh S., Ali A., Ullah A., Alenezi F., Gechev T., Hussain T., Ali F. 2023. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in Plant Science 14: 1158933. DOI: https://doi.org/10.3389/fpls.2023.1158933

Snell N. 2023. Modern medicines from plants. CRC Press: 51–62. DOI: https://doi.org/10.1201/9781003413004

Wang S., Wang Y., Chang Y., Zhao R., She Y. 2023. EBSE-YOLO: high precision recognition algorithm for small target foreign object detection. IEEE Access 11: 57951–57964. DOI: https://doi.org/10.1109/ACCESS.2023.3284062

Yan Y. 2022. Using the improved SSD algorithm to motion target detection and tracking. Computational Intelligence and Neuroscience 2022 (1): 1886964.