Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1, 668-681
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i1.4209

Detection of cocoa pod diseases using a hybrid feature extractor combining CNN and vision transformer with dual classifier

Kouassi Simeon KOUASSI[1*], Mamadou DIARRA[1], Kouassi Hilaire EDI[2], KOUA Brou Jean-Claude1[1]
[1]Laboratory of Mechanics and Computer Science, Félix Houphouët-Boigny University, Abidjan, Ivory Coast; simeon.kouassi@gmail.com (K.S.K.)
[2]Mathematics and Computer Science Laboratory, Nangui Abrogoua University, Abidjan, Ivory Coast; patoudiarra@gmail.com (M.D.) edi.hilaire@gmail.com (K.H.E.) k_brou@hotmail.com (K.B.J.C.)

Abstract: Ivory Coast is the world's leading cocoa producer, with a harvest of over two million tons in 2023. This dominance is severely threatened by several diseases, including Swollen Shoot, which was first detected in 1943. The epidemic, particularly severe in 2003, destroyed over 77,000 hectares of plantations. The prevalence of the disease has continued to rise, affecting new farms and causing substantial economic losses. In response to this situation, research is focusing on innovative solutions, such as artificial intelligence, to automate the detection of these diseases. In this study, a computer vision system was developed using a hybrid algorithm that combines convolutional neural networks and Transformers for feature extraction, along with dual classification through SVM and LightGBM for Swollen Shoot symptom detection. Our algorithm analyzes images of cocoa pods to identify symptoms such as reddish or black spots, brown or black lesions, or pod malformation. This model achieved an accuracy rate of 99.24%, surpassing other similar methods mentioned. Promising results for the classification of cocoa pod images, pave the way for practical solutions in the management of the Swollen Shoot pandemic. Through mobile applications or embedded systems, this technique will also contribute to the needs of early detection and intervention for precision agriculture.

Keywords: *CNN, Computer vision, Deep learning, Light GBM, Machine learning, SVM, Transformer, XGBOOST.*

1. Introduction

Côte d'Ivoire leads the global cocoa production market, producing around two million tons of beans in 2023 [1, 2]. Cocoa is a vital cash crop for the country's economy. However, this valuable income source faces a major challenge in plantation management: the detection and prevention of cocoa pod diseases. Diseases such as brown rot (caused by Phytophthora palmivora) and Cacao Swollen Shoot Virus (CSSV) present serious threats to cocoa production, thereby affecting the national economy, drastically reducing yields, and negatively impacting the livelihoods of farmers. Early and precise detection of these diseases is crucial to minimizing losses and improving harvest quality [3-5].

Since 2016, the spread of Swollen Shoot disease in Côte d'Ivoire has continued to raise concerns. Recent data indicate an increased presence in several new areas, affecting many other cocoa farmers. The disease's prevalence continues to grow, resulting in significant economic losses for farmers and jeopardizing the country's leading position in the global market [5, 6]. The symptoms of Swollen Shoot disease in cocoa trees are diverse and severely affect both the health of the trees and cocoa production.

1.1. Symptoms of Swollen Shoot Disease

The cocoa Swollen Shoot virus is semi-persistently transmitted by several species of mealybugs (Pseudococcidae, Homoptera) on cocoa plants [7, 8]. Infection occurs when the mealybugs acquire the

History: Received: 24 November 2024; Revised: 26 December 2024; Accepted: 6 January 2025; Published: 13 January 2025
* Correspondence: simeon.kouassi@gmail.com

virus from infected cocoa plants or other host plants and transmit it to healthy cocoa plants during feeding. The virus affects all parts of the cocoa plant. Symptoms of CSSV can be observed on the leaves, stems, roots, and cocoa pods [8]:

- In young leaves of particularly sensitive cocoa varieties, red vein banding may appear, which fades as the leaves mature.
- In mature leaves, a variety of symptoms may occur, such as yellowing along the veins, small and large speckles, spots, or streaks.
- A common symptom is chlorotic banding along the veins.
- Swelling of the stem (nodes, internodes, tips) and roots occurs due to abnormal cell proliferation.
- Distortion in the shape and color of the pods may also be observed.
- Infected cocoa pods become stunted, and cocoa production declines significantly.
- In severe cases, the disease can lead to the death of branches or even the entire tree.

1.2. Control Protocol for Swollen Shoot Disease

Several mitigation measures have been proposed to curb the spread of the virus, including:

- Identification and diagnosis of infected trees: This phase involves detecting Swollen Shoot symptoms in affected plants during inspections of contaminated areas. Technicians search for distinctive signs such as swellings and colored spots on the leaves.
- Removal of contaminated trees: After diagnosis, contaminated trees are removed to prevent further spread of the virus. This includes felling and destroying diseased trees to reduce the source of infection.
- Cross-protection with mild strains: This has been considered one of the alternative management strategies [9]. This biological control method involves inoculating plants with a mild (less virulent) strain of the virus. This mild strain infects the plant without causing severe symptoms, while preventing more virulent (and dangerous) strains from taking hold.
- Use of resistant plants (or replanting): This step involves using resistant or tolerant plant varieties to reduce the impact of the disease. Replanting is done at a minimum distance of 10 meters from old plantations to avoid cross-contamination.

By rigorously applying these protocols, the spread of Swollen Shoot can be controlled, protecting cocoa plantations and ensuring the sustainability of cocoa production in Ivory Coast. In addition to these measures, some programs integrate farmer awareness and training, as well as continuous monitoring and research to ensure the effectiveness of the implemented protocols.

1.3. Role of Machine Learning in Detecting Swollen Shoot

Early and accurate detection of these diseases is crucial for implementing effective management measures and minimizing losses. Traditionally, the identification of cocoa pod diseases has relied on visual inspections carried out by expert agronomists [10-12]. While this method is useful, it has significant limitations, such as the subjectivity of evaluations, the need for specialized knowledge, and the inability to comprehensively inspect large plantations. Moreover, early stages of infection may present subtle symptoms that are difficult to detect with the naked eye, delaying intervention and exacerbating damage. With the advent of artificial intelligence (AI) and machine learning technologies, new opportunities are emerging to automate and improve the process of detecting plant diseases. Convolutional Neural Networks (CNN) have demonstrated remarkable effectiveness in image recognition tasks, particularly in identifying plant diseases from visual data. Simultaneously, Vision Transformers (ViT) have recently gained popularity for their ability to capture complex spatial relationships and efficiently process information at different scales. These models leverage self-attention mechanisms that enable a more thorough contextual understanding of images, delivering superior

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

performance in various computer vision applications. Hybrid feature extractors, such as CNN and Vision Transformers (ViT), allow for an in-depth analysis of cocoa pod images, facilitating the identification of characteristic disease symptoms. Additionally, advanced classifiers like LightGBM (Light Gradient Boosting Machine) enhance the accuracy of predictions while being lightweight and fast to deploy in the field [13-15]. This study proposes an innovative approach for detecting cocoa pod diseases in Ivory Coast, with a particular emphasis on Swollen Shoot disease. We will examine emerging methods and technologies and discuss the opportunities offered by artificial intelligence to strengthen the resilience of cocoa plantations and support agricultural communities facing this persistent threat. The remainder of this document is structured as follows: Section II details the preprocessing of collected images and the feature extraction methods used. Section III presents the proposed methodology. Section IV discusses the experimental results and provides a comparison with existing algorithms applied to the same dataset. Finally, Section V concludes the study and explores future research directions.

2. Literature Review

Several studies have been conducted in the field of computer vision, particularly for disease detection in agriculture. In our study, we present some of these studies that are directly related to our investigation topic. Lomotey, et al. [15] used mobile technology and machine learning techniques based on deep Convolutional Neural Networks (CNN), to enhance the early detection and diagnosis of the two main diseases affecting cocoa production, namely swollen shoot and black pod disease. More specifically, they used a distributed mobile application developed to allow farmers to take a photo or video of cocoa pods, which the app automatically analyzes and detects the specific disease. The app then suggests the best treatment to undertake using an integrated information guide. The research analyzed 2,828 cocoa images distributed across three classes. It then built and trained four CNN models, such as: CentreNet ResNet50 V2, EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1 FPN [16]. The SSD MobileNet V2 model was the most generalized and fastest, with a detection confidence score of around 88.0%. Mamadou, et al. [17] used hybrid convolutional neural network approaches to identify diseases in cocoa pods, combining the MobileNetV2 architecture with sophisticated classification algorithms such as logistic regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), XGBoost, and Random Forest. They performed the work in two phases; first, each algorithm was individually evaluated, then their performances were measured when combined with MobileNetV2. This hybrid approach enhanced MobileNetV2's capabilities, achieving accuracy rates ranging from 72.4% to 86.04%, demonstrating the effectiveness of combining MobileNetV2 feature extraction with the classification skills of other algorithms. Ayikpa, et al. [18] conducted a detailed study to assess the effectiveness of feature extractors and the influence of different color spaces on these extractors for classifying cocoa pod maturity from images. They used several similarity measures in conjunction with these extractors to categorize images based on their maturity level. The study analyzed the performance of extractors based on CNNs and the Gray Level Co-occurrence Matrix (GLCM) in various color spaces, including RGB, HSV, Lab, and Luv. The results revealed that the Lab color space, combined with GLCM and the chi-square distance similarity measure, offered the best performance, achieving an accuracy of 99.60%.

IEEE Xplore Abstract Record [19] proposed using convolutional neural networks to classify cocoa diseases, with a particular focus on the severe agricultural and economic consequences of black pod rot and cocoa pod borers. They worked with a set of 4,390 images and evaluated five CNN architectures: a custom CNN, VGG-16, EfficientNetB0, ResNet50, and LeNet-5 for their ability to accurately identify the presence of these diseases. The custom CNN model stood out as the best performer, achieving an accuracy, recall, and precision of 91.79%, an F1 score of 82.08%, a sensitivity of 96.69%, and a specificity of 98.40%, demonstrating high efficiency in distinguishing healthy from diseased plants.

Coulibaly, et al. [20] proposed an effective system for recognizing Swollen Shoot disease symptoms by extracting features from cocoa pods. Their approach relied on using deep convolutional neural networks (CNNs) to improve the diagnosis of infected plants. The results of their study demonstrated

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

that a deep convolutional neural network could achieve an accuracy rate of 84% on a supervised learning dataset. Ayikpa, et al. [21] used several approaches for classifying and recognizing coffee leaf diseases, based on both traditional machine learning and deep learning methods. The evaluation of these approaches showed effectiveness, with 100% accuracy for traditional learning methods such as SVM and Random Forest, and for deep learning methods such as MobileNet and custom CNNs.

Bueno, et al. [22] studied a technique to determine the maturity degree of cocoa based on the acoustic sound of cocoa beans. Then, using the acoustic sound of cocoa pods, their approach extracted recognizable features for training purposes and applied convolutional neural networks (CNN) to classify the sound of cocoa. The experimental model gave a classification accuracy of 97.46% for the system on the maturity question of unharvested cocoa pods. Kumi, et al. [16] developed a smartphone application based on deep learning for detecting the Swollen Shoot disease and black pod rot. This mobile application, designed with integrated machine learning techniques, allows cocoa producers to take a photo of the cocoa pod and upload it for diagnosis, which takes place on a cloud service. They used four built and trained CNN models, and at the end of their study, they obtained an accuracy of over 80% with the SSD MobileNet V2 model. Vera, et al. [23] presented a deep learning-based model applied to the identification of cocoa pod diseases, focusing on "moniliasis" and "black rot" using the EfficientDet-Lite4 model [23] an efficient and lightweight model for object detection. A dataset, comprising images of healthy and diseased cocoa pods, was used to train the model to detect and locate disease manifestations with considerable accuracy. Moreover, the model's features were integrated into a native Android mobile application with a user-friendly interface, enabling young or inexperienced farmers to quickly and accurately identify the health status of cocoa pods. Rodriguez, et al. [24] proposed a machine learning approach to identify cocoa tree diseases (Theobroma cacao L.) and prevent crop loss, as farmers lack immediate tools to detect diseases on time. They used image processing and analysis techniques such as HoG (Histogram of Oriented Gradients) [25] LBP (Local Binary Pattern) [26] and the SVM (Support Vector Machine) [27, 28] classification algorithm to determine whether the cocoa tree is infected by a disease. The results obtained show that applying SVM, Random Forest, and Artificial Neural Networks (ANN) with the feature vectors.

3. Methodology

This section outlines the proposed methodology for our study, including the dataset collection, data preprocessing approaches, the materials used, the algorithms description, the deep learning methods, and the models proposed.

3.1. Dataset Description

The data used in this study comes from a collection of cacao pod images of the Forastero variety, specifically the Amelonado sub-variety and Trinitario variety. The images were collected from three plantations located in Côte d'Ivoire: the experimental plantation of the National Agronomic Research Center (CNRA). Located in Bouaflé, in the central-western region of Côte d'Ivoire (Forastero) and a local farmer's plantation in the same area (Forastero group). The last is a local farmer's plantation in Divo (Forastero and Trinitario group), located in the southern part of Côte d'Ivoire, a region known for high cocoa production. This dataset contains 3,500 images with a resolution of 2456px by 3275px, divided as follows: 1,200 images of healthy pods and 2,300 images of infected pods showing visible symptoms. The photos were taken in real-world conditions, under open skies, between 10 a.m. and 4 p.m., with ambient temperatures ranging from 28°C to 32°C, solar irradiance between 200 to 500 W/m², and relative humidity between 70% and 90%. The camera distances varied from one to three meters. Figures 1 and 2 respectively show examples of healthy and symptomatic cacao pods.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

Figure 1.
Example of healthy cacao pods.

Figure 2.
Example of cacao pods showing symptoms of SWOLLEN SHOOT.

3.2. Database Preprocessing

Data preprocessing is an essential step in any image classification process, significantly contributing to the performance and efficiency of the models. Thanks to recent innovations, data preprocessing has evolved to incorporate more advanced and sophisticated techniques, providing numerous benefits. Modern innovations and image preprocessing techniques help with:

- Noise and artifact removal,
- Color adjustment,
- Model accuracy improvement,
- Image size standardization,
- Data augmentation,
- Reduced training time,
- Focus on relevant features.

For the preprocessing of our data (images), we employed the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm [29-31].

3.2.1. The CLAHE Algorithm (Contrast Limited Adaptive Histogram Equalization)

The CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm is an image processing method designed to enhance local contrast while reducing the amplified noise often introduced by global histogram equalization [32].

The CLAHE algorithm involves several key steps to improve image contrast:

- **Image Division**: The image is segmented into blocks (256x256 pixels) to process each region locally while preserving its unique features.
- **Histogram Equalization**: Each block undergoes uniform redistribution of brightness, thereby enhancing local contrast.
- **Contrast Limitation**: A threshold is applied to prevent excessive noise amplification in homogeneous areas, ensuring high visual quality.
- **Interpolation**: The borders between blocks are smoothed to ensure seamless transitions without visible artifacts.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

After processing all the blocks, the resulting image shows improved local contrast without the typical artifacts associated with global histogram equalization. Below are some images resulting from the application of this preprocessing algorithm.

Figure 3.
Example of Preprocessed Images.

3.2.2. Data Augmentation

To enhance the model's generalization ability by exposing it to different perspectives of the original image, we applied data augmentation by rotating the image at three angles: 90°, 180°, and 270°, followed by flipping each of these images. This technique allows us to increase the size of our dataset by a factor of eight. The main advantage of this method is that it diversifies the dataset while making the model more resilient to noise. As a result, it improves both the precision and sensitivity of the model by optimizing feature detection and reducing bias.

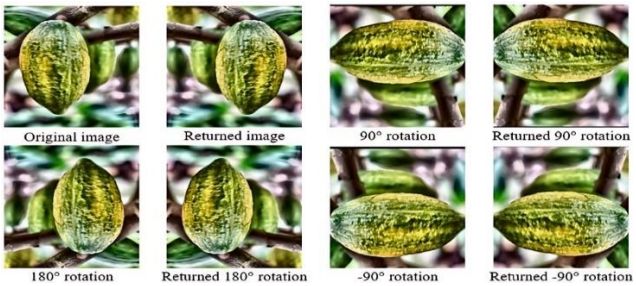

Figure 4.
Example of dataset augmentation.

3.2.3. Equipment Used

For our project, we chose Python due to its ease of use, flexibility, and the extensive range of available libraries. Additionally, it is compatible with most online platforms. As for hardware, we used an HP all-in-one desktop PC, equipped with an Intel (R) CORE I7-12700T processor running at 2.4 GHz, 24 GB of RAM, a 1 TB SSD, and an NVIDIA Quadro P400 graphics card. For testing purposes, we also utilized the online platform Kaggle, which offers a working environment with 16 GB of disk space, 16 GB of RAM for the CPU, and 16 GB of RAM for the GPU.

3.2.4. Feature Extraction Algorithm

Ultimately, our work will involve automatic detection of the Swollen Shoot symptoms listed above. This is a computer vision task based essentially on machine learning. In this study, we used in parallel a CNN algorithm and a Transformers vision algorithm for feature extraction, and the SVM and LightGBM algorithms for classification.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

3.2.5. Convolutional Neural Networks (CNN) Algorithms

Convolutional Neural Networks (CNNs) are a type of artificial neural network specifically designed to process grid-structured data, such as images. Unlike traditional neural networks, CNNs leverage the spatial relationships between pixels in an image, making them particularly effective for computer vision tasks.

A CNN consists of several layers, each playing a distinct role in data processing.

Convolution is the fundamental operation in CNNs. It involves sliding a filter (or kernel) over the input image to compute a dot product between the filter and sub-regions of the image. The result of this operation is a feature map, which captures patterns present in the image.

$$S(i,j) = \sum_m \sum_n I(i+m, j+n) * K(m,n) \quad (1)$$

Where S (I, j) is the pixel of the feature map at position (I, j), I is the input image, K is the convolution kernel (filter), (i,j) represents the position in the image

After the convolution, a non-linear activation function, such as the ReLU (Rectified Linear Unit) function, is applied to introduce non-linearity. This helps the network learn patterns that are more complex.

$$f(x) = \max(0, x) \quad (2)$$

Where x the input value.

Pooling (often max-pooling) reduces the dimensionality of the feature maps while retaining important information. This helps decrease the size of the data being processed, making the model more robust to small translations and deformations in the image.

$$P(i,j) = max(S(i{:}\, i+f, j{:}\, j+f)) \quad (3)$$

Where P(i ,j) is the pooling result for the region (i:i+f,j:j+f) of the feature map S, with f representing the size of the pooling filter.

After several layers of convolution and pooling, the extracted features are flattened (i.e., transformed into a vector) and passed through fully connected layers. These layers operate like those in a traditional neural network, combining the extracted features to make predictions.

The final layer is often a softmax layer in classification tasks, producing a probability for each possible class. In an image classification task, the CNN takes an image as input and, after going through the aforementioned steps, outputs a probability for each category. The model is then trained using backpropagation, where the prediction error, usually measured by a loss function like cross-entropy, is minimized by adjusting the weights of the filters through gradient descent.

CNNs are widely used for tasks such as object recognition, face detection, pathology classification, and many other computer vision applications.

As part of this work, we will use the CNN model DenseNet121.

3.2.6. DenseNet121 Model

DenseNet121 (Dense Convolutional Network) is a convolutional neural network (CNN) model that stands out for how it connects layers together. Unlike traditional CNNs, where each layer is only connected to the next, DenseNet connects each layer to all subsequent layers. This allows for better reuse of learned features, reduces the number of parameters, and facilitates the training of deep networks. DenseNet121 is a version with 121 layers. Its main advantages include:

- Efficient gradient propagation: avoiding the vanishing gradient problem.
- Reduced overfitting: through more information that is effective sharing.
- Simplicity: compared to other deep architectures, as there is no need to recalculate already extracted features.

DenseNet121 proves to be very effective for image classification tasks, segmentation, and other computer vision applications.

Its ability to automatically extract relevant features from image data makes it a powerful tool in the field of artificial intelligence.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

3.2.7. Vision Transformers Algorithms

Vision Transformers (ViT) are Transformer architectures initially designed for natural language processing (NLP) and are particularly suitable for computer vision tasks. These models apply the Transformer architecture to images. Instead of analyzing sequences of words as in NLP, ViTs process sequences of image patches. A patch is a subset of an image, typically of fixed size, which is treated as an element of the input sequence for the Transformer model.

The process of processing an image with a Vision Transformer can be described in several key steps:

- Dividing the Image into Patches: The input image is divided into a set of square patches, each being of fixed size (e.g., 16x16 pixels). Each patch is flattened into a vector.
- Encoding the Patches: Each patch is then transformed into a higher-dimensional vector using a linear embedding. Additionally, positional encoding is added to retain the spatial information of the patches in the original image.
- Passing through Transformer Layers: The encoded patch vectors are then processed by a Transformer model, which uses attention mechanisms to capture the relationships between different parts of the image. The Transformer model consists of multiple layers of self-attention and feed-forward.
- Classification: After passing through all the Transformer layers, the classification vector is extracted (often called the class token), which is then passed through a fully connected classification layer to produce the final predictions.

The classification task with a Vision Transformer involves assigning a class label to the input image. The model learns a prediction function f(x).

$$f(x) = y \quad (1)$$

Where x is the input image and y is the class prediction.

The output of the final layer of the Transformer, after applying the classification layer, is a probability distribution over the different classes. The softmax function is often used to obtain class probabilities:

$$y = \text{softmax}(Wz + b) \quad (2)$$

Where z is the extracted classification vector, W is a weight matrix, and b is a bias vector.

3.2.8. Mixed CNN-ViT Algorithms

Mixed CNN-ViT models combine the advantages of Convolutional Neural Networks (CNN) with those of Vision Transformers (ViT), to handle complex computer vision tasks. These models seek to leverage the ability of CNNs to extract local features and the efficiency of ViTs to capture global relationships and model long-term dependencies in images.

As part of our work, we implemented a hybrid architecture that combines the advantages of CNNs and transformer models. Based on the performance observed during our experimental tests, we opted for the DenseNet121 model without the fully connected layer. DenseNet121 CNN proved to be highly effective for extracting local features from images, using convolutional filters that detect edges, textures, or shapes. The DenseNet model is particularly good at capturing fine details in images. In parallel with this model, we integrated a Transformer algorithm, which we adapted to our dataset.

Thanks to their attention mechanism, the Transformer model excels at capturing long-range dependencies. They allow for modeling global relationships in images by analyzing how each part of the image is related to others, regardless of spatial distance.

This hybridization allowed us to achieve both fine-grained, localized understanding of images, while also capturing global dependencies. The CNN-Transformer hybrid, combined with dual classification, provided us with the best prediction scores and significantly improved the performance of our model.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

3.2.9. Classification algorithm

3.2.9.1. The XGBoost Classifier

The XGBoost (Extreme Gradient Boosting) classifier is an optimized implementation of the gradient boosting algorithm designed to be highly performant and efficient in terms of time and memory. It is widely used in the Machine Learning community for its ability to produce high-precision models while effectively managing bias and variance.

The XGBoost algorithm is based on the idea of sequentially building an ensemble of weak models (often decision trees). Each new model attempts to correct the errors made by previous models by adjusting the weights of observations according to the residual error, and then minimizing a loss function.

XGBoost is widely used for several reasons:

- Speed: The algorithm's performance on large datasets.
- Handling Missing Values: XGBoost can manage missing values, which is an advantage over other algorithms.
- Regularization: It includes mechanisms such as L1 and L2 regularization to prevent overfitting.
- Parallelism: It exploits parallelism to accelerate the training process.
- Customization: It allows for the customization of loss functions and other aspects of the algorithm.

3.2.10. The SVM Classifier

The SVM (Support Vector Machine) is a highly effective classification algorithm, particularly useful for solving both linear and non-linear classification problems. Its basic principle is to find an optimal hyperplane that separates the data into two distinct classes while maximizing the margin between this hyperplane and the closest points from each class (called support vectors).

SVMs are efficient in high-dimensional spaces and can handle cases where the data is not linearly separable by using "kernels" to project the data into a higher-dimensional space. They stand out for their ability to generalize well to new data, especially in contexts where the number of dimensions exceeds the number of samples. In terms of performance, SVMs are robust, minimize overfitting, and can handle complex datasets.

3.2.11. The LightGBM Classifier

LightGBM (Light Gradient Boosting Machine) is a machine-learning algorithm based on the boosting technique. It is designed to be efficient, fast, and capable of handling large datasets while maintaining high accuracy.

LightGBM is based on the concept of gradient boosting, where multiple decision trees are trained sequentially. Each new tree attempts to correct the errors of previous trees by focusing on misclassified samples.

It is designed to be faster than other boosting algorithms, such as XGBoost, when processing large datasets. It uses techniques such as leaf-wise tree growth rather than level-wise, allowing for better optimization.

Particularly suited for handling large volumes of data, LightGBM employs a data loading technique called GOSS (Gradient-based One-Side Sampling), which reduces the number of samples while maintaining accuracy. The algorithm directly manages categorical variables without needing to transform them into numerical variables, or use the techniques like one-hot encoding.

LightGBM is widely used for tasks such as classification, regression, ranking, and even time series prediction.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

3.2.12. Algorithms Implementation

3.2.12.1. Architecture

Our algorithm follows the general architecture of hybrid networks. See Figure 5.

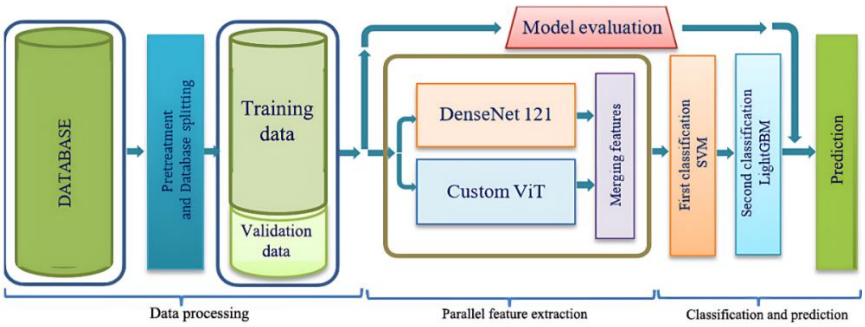

Figure 5.
Methodology flowchart.

3.2.13. Hyper Parameters

We implemented this algorithm using the Python programming language with TensorFlow libraries, notably Keras. In the implementation, we used the following hyper parameters:

- n_estimators: Number of trees to build.
- learning_rate: Learning rate for adjusting weights after each tree.
- max_depth: Maximum depth of the trees, which controls model complexity.
- subsample: Fraction of the sample used to build each tree.
- colsample_bytree: Fraction of features used to build each tree.
- num_leaves: Maximum number of leaves in each tree.
- min_data_in_leaf: Minimum number of samples needed in a leaf.

3.2.14. Algorithms From the Literature

To compare our results, we selected several algorithms from the literature for their proven performance in detecting and classifying agricultural pathologies. These algorithms were also implemented and tested in our environment using the same dataset.

3.2.15. DeiT

DeiT (Data-efficient Image Transformer) is a transformer-based model designed for image classification. It improves training efficiency by using a distillation token, leveraging both labeled data and a teacher model for better performance. Its lightweight architecture makes it suitable for various computer vision tasks.

3.2.16. ConViT

The ConViT (Convolutional Vision Transformer) combines convolutional neural networks (CNNs) with transformers for image classification, leveraging the local feature extraction of CNNs and the global attention mechanism of transformers. It balances efficiency and accuracy, making it suitable for various vision tasks.

4. Results

To evaluate the performance of extractors and classifiers in detecting swollen shoot symptoms based on cocoa pod images, we organized our tests into two scenarios.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

4.1. Test Descriptions

4.2.1. Test Scenario 1: Experience 1

Detection of swollen shoot symptoms based on cocoa pod images using feature extractors from convolutional neural network architectures (CNN) DenseNet121, Vision Transformer (ViT) DeiT, and a mixed CNN-ViT architecture ConViT. For this first experience, we used fully connected RNN layers as classifiers.

4.2.2. Test Scenario 1: Experience 2

In this second test, we again detected swollen shoot symptoms based on the same dataset using feature extractors from the three architectures used in experience 1 (DenseNet121, DeiT, and ConViT), but with the SVM classifier.

4.2.3. Test Scenario 1: Experience 3

In this third test, we again detected swollen shoot symptoms based on the same dataset using feature extractors from the three architectures used in experience 1 (DenseNet121, DeiT, and ConViT), but with the XGBoost classifier.

4.2.4. Test Scenario 1: Experience 4

In this final scenario, we used the same feature extractors with the same dataset in the same environment. However, unlike the previous experience, we used the LightGBM classifier.

4.3. Test Scenario 2

In this second scenario, we used our own architecture for feature extraction, which runs the DenseNet121 model in parallel with a custom ViT model. Then, these features are classified using two classifiers: SVM (Support Vector Machine) and LightGBM, as described in section 3.4.4.

4.4. Evaluation Metrics

To assess the performance of the models in our study, we will use various evaluation usual metrics, which are:

- Accuracy,
- Mean Squared Error,
- Recall,
- F1 score,
- Matthews Correlation Coefficient.

4.5. Results of the Scenarios

Tables below summarize the results of the tests conducted, as presented above, according to the previously mentioned metrics for Scenario 1.

4.6. Summary of Scenario 1 Tests Results

This section presents all the experimental results obtained during the work conducted, as well as a comparative study of the results obtained from the DenseNet121, DeiT, and ConViT algorithms combined with fully connected RNN, SVM, XGBoost, and LightGBM classifiers.

To highlight the differences and similarities in the implementation of the two scenarios, we will use the following common metrics.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

Table 1.
Performance measures for deep learning models.

	Metrics	Scenario 1: Exp 1 RNN fully connected classifier	Scenario 1: Exp 2 SVM classifier	Scenario 1 : Exp 3 The XGBoost classifier	Scenario 1 : Exp 4 The LightGBM classifier
DenseNet121	Accuracy	87.26%	98.58%	91.04%	93.87%
	Precision	85.31%	99.22%	88.28%	94.03%
	Recall	95.31%	98.46%	98.46%	96.18%
	F1 Score	90.04%	98.84%	93.09%	95.09%
	MSE	0.1092	1.42%	0.8960	0.5190
	MCC	0.7658	97.03%	0.8142	0.8909
DeiT	Accuracy	92.45%	97.17%	94.34%	94.81%
	Precision	95.97%	98.44%	97.58%	97.62%
	Recall	91.54%	96.92%	93.08%	93.89%
	F1 Score	93.70%	97.67%	95.28%	95.72%
	MSE	0.2083	2.83%	0.566	0.519
	MCC	0.6060	94.08%	0.8838	0.8929
ConViT	Accuracy	95.28%	98.11%	96.23%	98.11%
	Precision	96.15%	97.73%	96.92%	98.46%
	Recall	96.15%	99.23%	96.92%	98.46%
	F1 Score	96.15%	98.47%	96.92%	98.46%
	MSE	0.2462	1.89%	0.3771	0.1892
	MCC	0.9146	96.02%	0.9205	0.9602

4.7. Scenario 2 Tests Results

This section presents the experimental results obtained during the tests of our second scenario. It highlights the performance of the algorithms tested in scenario 1 and our own architecture for feature extraction.

4.8. Summary of Results for Both Scenarios

Table 2.
Performance measures for hybrid CNN- ViT algorithm models.

Scenario 2: Hybrid CNN- ViT algorithm with dual classification (SVM and LightGBM)					
Accuracy	**Precision**	**Recall**	**F1 Score**	**MSE**	**MCC**
99.53%	99.24%	100%	99.62%	0.0047	0.9901

5. Discussion

Swollen Shoot, a disease affecting cocoa trees, poses a significant threat to agriculture and reduces productivity, leading to substantial economic losses. Numerous studies have been conducted to detect plant diseases, including Swollen Shoot, using machine learning techniques. The results of our study reveal that the applied methods are highly effective in classifying and recognizing Swollen Shoot symptoms from images of cocoa pods. The deep learning algorithms used in our work demonstrated high efficiency, with accuracy rates ranging from 85.31% to 99.22%. In Table 1, which presents the metrics values collected during the validation and testing phases of the deep learning models, we obtained accuracy values ranging from 87.26% to 98.58% and recall values from 95.31% to 99.23%, confirming the overall predictive performance of the models and their strong ability to correctly classify positive cases. Precision values ranged from 85.31% to 99.22%, and F1-scores ranged from 90.04% to 98.84%, confirming the models' balanced ability to identify positive instances while maintaining a low false-positive rate. SVM and LightGBM classifiers proved to be much more efficient, accurate, and faster for classification tasks compared to XGBoost classifiers and recurrent neural networks (RNNs). The hybrid model (resulting from the combination of CNN-ViT models and SVM-LightGBM classifiers) stood out from the others in terms of overall performance, achieving the highest accuracy rates in the second scenario studied (with accuracy rates from 99.24% to 99.53%) and a recall of 100%, thus ensuring more precise classification of positive and negative cases. Its F1-score of 99.62% reflects a perfect

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1: 668-681, 2025
DOI: 10.55214/25768484.v9i1.4209

balance between precision and recall, while its MCC of 99.01% indicates excellent classification quality. With an accuracy of 99.53%, this model stands out for its precision in classifying instances.

6. Conclusion

In this study, we explored several deep learning techniques based on recurrent neural networks (RNN), as well as SVM, XGBoost, LightGBM, and hybrid classifiers (combination of SVM and LightGBM classifiers), for the classification and recognition of Swollen Shoot disease on cocoa pods. The evaluation of these methods confirmed the overall effectiveness of the five classifiers, with accuracies ranging from 85.31% to 99.22%. The hybrid classifier (combination of SVM and LightGBM classifiers) achieved the highest accuracy rate, standing out from the others with a remarkable accuracy of 99.53% and a low loss (0.47%) in the studied scenarios, demonstrating its robustness. In the future, we plan to implement a resilient approach in uncontrolled environments by integrating our method with a segmentation technique to better target the areas of cocoa pods affected by Swollen Shoot.

Our dataset was composed of two cocoa sub-varieties available as part of the Swollen Shoot control project. In our future work, we will expand this dataset to include other varieties cultivated in Brazil and globally. We also plan to leverage the advantages of new feature extraction and classification architectures to further enhance these results.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgments:

We would like to express our sincere thanks and gratitude to all the individuals who supported us throughout this study. We also extend our appreciation to the person in charge at Felix Houphouët-Boigny University.

References

[1] R. Connection Ivoirienne, "Rapid rise of cocoa processing in Ivory Coast, Ivorian Connection," Retrieved: https://connectionivoirienne.net/03/15/ascension-fulgurante-de-la-transformation-de-cacao-en-cote-divoire/. [Accessed 2023.

[2] -. Côte d'Ivoire: Agence Afrique, "Ivory Coast: 1,500 FCFA per kilogram for the 2023-2024 intermediate cocoa campaign (Official)," Retrieved: https://www.agenceafrique.com/52037-cote-divoire-pour-la-campagne-intermediaire-du-cacao-officiel.html. [Accessed n.d.

[3] Y. Wilfried Junior, K. Séka, and A. L. F. H. Koffi, "Diversity of Phytophthora palmivora and Trichoderma sp. strains in cocoa orchards in 3 regions of Côte d'Ivoire," *Afrique Science*, vol. 24, no. 5, pp. 69-79, 2024.

[4] T. K. Guessan-Bi, K. D. Kra, K. É. Kwadjo, K. L. Kouame, and M. Doumbia, "Farmers' practices for the orchard's maintenance and post-harvest treatment of cocoa in infiltrated classified and unclassified zone of Méagui (South-West, Côte d'Ivoire)," *Journal of Agricultural Chemistry and Environment*, vol. 12, no. 3, p. 275-295, 2023.

[5] F. Ruf, M. Salvan, J. Kouamé, and T. Duplan, "Who are the cocoa farmers of Ivory Coast?," *Papiers de Recherche*, pp. 1-111, 2020.

[6] O. Domfeh *et al.*, "Evaluation of mild strain cross protection in cacao–further evidence of the protective potential of cacao swollen shoot virus strain N1 against the New Juabeng (1A) isolate under field conditions," *Australasian Plant Pathology*, vol. 50, pp. 329-340, 2021. https://doi.org/10.1007/s13313-021-00777-1

[7] T. CACAO, "CACAO swollen shoot virus," Doctoral Dissertation, University of the West of England, Bristol, 2024.

[8] F. B. Agusto, M. C. Leite, F. Owusu-Ansah, O. Domfeh, N. Hritonenko, and B. Chen-Charpentier, "Cacao sustainability: The case of cacao swollen-shoot virus co-infection," *Plos One*, vol. 19, no. 3, p. e0294579, 2024. https://doi.org/10.1371/journal.pone.0294579

[9] O. Domfeh, G. Ameyaw, H. Dzahini-Obiatey, and L. del Río Mendoza, "Spatiotemporal spread of cacao swollen shoot virus severe strain 1A in mixed hybrid cacao pre-inoculated with mild strain N1," *Plant Disease*, vol. 103, no. 12, pp. 3244-3250, 2019. https://doi.org/10.1094/pdis-12-18-2175-re

[10] E. A. Gyamera, O. Domfeh, and G. A. Ameyaw, "Cacao swollen shoot viruses in Ghana," *Plant Disease*, vol. 107, no. 5, pp. 1261-1278, 2023. https://doi.org/10.1094/pdis-10-22-2412-fe

[11] X.-S. Zhang and J. Holt, "Mathematical models of cross protection in the epidemiology of plant-virus diseases," *Phytopathology*, vol. 91, no. 10, pp. 924-934, 2001. https://doi.org/10.1094/phyto.2001.91.10.924

[12] K. N. D. Nobert, C. Klotioloma, D. Mamadou, and N. G. K. Francois, "Host plants identification of mealybugs main species, vectors of the swollen shoot virus in the counties of Abengourou, Bouaflé and Divo," *Journal of Applied Biosciences*, vol. 192, pp. 20366-20377, 2023.

[13] G. A. Ameyaw, O. Domfeh, and E. Gyamera, "Epidemiology and diagnostics of cacao swollen shoot disease in Ghana: Past research achievements and knowledge gaps to guide future research," *Viruses*, vol. 16, no. 1, p. 43, 2023. https://doi.org/10.3390/v16010043

[14] J. K. Appati, "Performance and applicability of transfer learners for cocoa swollen shoot detection," *International Journal of Technology Diffusion*, vol. 12, no. 2, pp. 68-77, 2021. https://doi.org/10.4018/ijtd.2021040105

[15] 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, pp. 92-119, 2024. https://doi.org/10.1504/ijsami.2024.10059217

[16] 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," *Procedia Computer Science*, vol. 203, pp. 87-94, 2022. https://doi.org/10.1016/j.procs.2022.07.013

[17] D. Mamadou, J. A. Kacoutchy, 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, pp. 1-9, 2023. https://doi.org/10.14569/ijacsa.2023.0140937

[18] K. J. Ayikpa, D. Mamadou, P. Gouton, and K. J. Adou, "Classification of cocoa pod maturity using similarity tools on an image database: Comparison of feature extractors and color spaces," *Data*, vol. 8, no. 6, p. 99, 2023. https://doi.org/10.3390/data8060099

[19] IEEE Xplore Abstract Record, "IEEE Xplore abstract record," Retrieved: https://ieeexplore.ieee.org/document/9322689. [Accessed 2024.

[20] 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, p. 166, 2020. https://doi.org/10.4236/eng.2020.123014

[21] K. J. Ayikpa, D. Mamadou, P. Gouton, and K. J. Adou, "Experimental evaluation of coffee leaf disease classification and recognition based on machine learning and deep learning algorithms," *Journal of Computer Science*, vol. 18, no. 12, pp. 1201-1212, 2022. https://doi.org/10.3844/jcssp.2022.1201.1212

[22] G. E. Bueno, K. A. Valenzuela, and E. R. Arboleda, "Maturity classification of cacao through spectrogram and convolutional neural network," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 3, pp. 228-233, 2020. https://doi.org/10.14710/jtsiskom.2020.13733

[23] D. B. Vera, B. Oviedo, W. C. Casanova, and C. Zambrano-Vega, "Deep learning-based computational model for disease identification in cocoa pods (Theobroma cacao L.)," *arXiv preprint arXiv:2401.01247*, 2024.

[24] C. Rodriguez, O. Alfaro, P. Paredes, D. Esenarro, and F. Hilario, "Machine learning techniques in the detection of cocoa (Theobroma cacao L.) diseases," *Annals of the Romanian Society for Cell Biology*, pp. 7732-7741, 2021.

[25] D. A. Forsyth and J. Ponce, *Computer vision: A modern approach*. Prentice Hall Professional Technical Reference, 2002.

[26] I. Riaz, A. N. Ali, and H. Ibrahim, "Circular shift combination local binary pattern (CSC-LBP) method for dorsal finger crease classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 8, p. 101667, 2023. https://doi.org/10.1016/j.jksuci.2023.101667

[27] H. Noria and A. Aniya, "Application of support vector machines (SVM) method for handwritten digit recognition," Doctoral Dissertation, Université Mouloud Mammeri, 2011.

[28] S. Bouillant, J. Mitéran, F. Yang, and M. Paindavoine, "Real-time defect detection on objects with complex geometry: Study by SVM and hyperrectangles," Retrieved: https://core.ac.uk/download/pdf/15494840.pdf. [Accessed 2003.

[29] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, 2017. https://doi.org/10.1109/TIP.2017.2662206

[30] Y. R. Haddadi, B. Mansouri, and F. Z. I. Khodja, "A novel medical image enhancement algorithm based on CLAHE and pelican optimization," *Multimedia Tools and Applications*, pp. 1-20, 2024. https://doi.org/10.1007/s11042-024-19070-6

[31] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, pp. 35-44, 2004. https://doi.org/10.1023/b:vlsi.0000028532.53893.82

[32] K. C. Paul and S. Aslan, "An improved real-time face recognition system at low resolution based on local binary pattern histogram algorithm and CLAHE," *arXiv preprint arXiv:2104.07234*, 2021. https://doi.org/10.4236/opj.2021.114005