Deep learning approach using capsule networks for cocoa plant disease detection from 3D images

Benoît Martin Azanguezet Quimatio[1] and Aurelle Joelle Nguka Ngoumgang[2]

[1] University of Dschang, Dschang, Cameroon
aurellengoumgang@gmail.com
[2] University of Dschang, Dschang, Cameroon
Azanguezet@gmail.com

Abstract. In response to the challenges posed by traditional convolutional neural networks in effectively capturing anomalies and deformations in cocoa pods, this thesis proposes a method for classifying images of diseased cocoa pods by integrating capsule neural networks (CapsNets) with a self-attention mechanism. By optimizing the CapsNet architecture, this model preserves the hierarchical and spatial information of the cocoa pods, while the self-attention mechanism enhances the model's focus on crucial features, thereby improving the accuracy of disease detection. Our approach also extends the capability of CapsNets to work with 3D images. The methodology includes collecting 3D images of cocoa pods, preprocessing this data by segmenting point clouds, and applying the enhanced CapsNets. Trained and tested on datasets of diseased cocoa pod images, the proposed model demonstrated an ability to accurately recognize anomalies, achieving a precision rate of 95%. Additionally, it achieved an F1 Score and recall of 98%, which is a significant improvement compared to the state-of-the-art performances of 83.05% for both the F1 Score and recall. Although our model shows a slightly lower precision than the state-of-the-art (98%), its performance in terms of F1 Score and recall indicates a better balance between precision and recall, thus highlighting its robustness in classifying cocoa pod diseases. This innovative approach, which had not been explored before, represents a promising solution for detecting diseased cocoa pods in the agricultural sector.

Keywords: Capsule Networks (CapsNets) , Self-attention mechanism , 3D and 2D images , Point cloud segmentation , Detection of cocoa pod diseases.

1 Introduction

The cacao tree (*Theobroma cacao* L.), belonging to the Malvaceae family and originating from the tropical forests of Central America, is a tropical plant cultivated for its fruit, cocoa, which is essential for chocolate production [1]. Today, this species is cultivated in many tropical countries, notably in Africa and Latin America. As a vital component of the global agricultural economy, especially in developing countries in tropical regions, the cacao tree is central to the livelihoods of millions of small farmers and agricultural workers. The cocoa sector significantly contributes to the global economy with approximately 8.6 billion US dollars in annual exports and supports 40 to 50 million people who depend on this crop for their income (ICCO, 2017). Cocoa production is vital not only for the economies of producing countries but also for downstream industries such as the chocolate industry, which used 43% of the world's production in 2017, representing a market value of 106.19 billion US dollars and is anticipated to grow to 189.89 billion US dollars by 2026. However, cocoa production is vulnerable to diseases that can compromise both the quality and quantity of the yield. Faced with this vulnerability, it is imperative to develop more efficient and precise detection methods. Traditional approaches based on visual observation are often insufficient, particularly for detecting subtle or early symptoms of diseases, leading to delayed interventions and excessive pesticide use. In a context where environmental concerns are increasing, it is essential to develop innovative solutions that minimize ecological impact while optimizing crop management. Convolutional neural networks (CNNs) have been a notable advancement in automating plant disease detection due to their ability to process two-dimensional images. However, they encounter significant limitations, notably the loss of hierarchical spatial information due to their pooling structure [2]. To overcome these limitations, capsule neural networks (CapsNets) have emerged as a promising solution. Unlike CNNs, CapsNets preserve spatial information by using activation vectors to represent not only the existence but also the spatial properties of entities in the image [2]. This allows CapsNets to better understand the scene by capturing the spatial relationship between parts and the whole, which is crucial for identifying cocoa tree diseases. However, despite these advances, CapsNets still face challenges, particularly their limited application to three-dimensional (3D) data. Current research shows that while CapsNets can theoretically process 3D data, existing methods are not robust enough to handle the complexity and variability of spatial data encountered in cocoa tree diseases [3] , [4] , [5] , [6]. The importance of overcoming these limitations is underscored by the crucial role that early and accurate disease detection plays in the sustainable management of cocoa crops. The subtle or spatially complex symptoms of cocoa tree diseases require detailed and multidimensional analysis for reliable diagnosis [3]. To address this issue, this research aims to bridge the gap between the theoretical capabilities of CapsNets and their practical application in 3D detection of cocoa tree diseases by optimizing the routing algorithm and integrating the processing of 3D point clouds [5]. Therefore, the problem to be addressed is: How to detect all the subtle modifications present on cocoa pods, which is essential for an accurate diagnosis of cocoa tree diseases?

2 Related work

The importance of early detection of plant diseases, especially in cocoa trees (Theobroma cacao), cannot be underestimated. These pathologies can lead to yield losses, economic damage, and significant environmental consequences. Faced with this challenge, numerous researchers have explored various approaches to monitor and identify these diseases, ranging from traditional methods to technological advancements. To begin, the study conducted by Awudzi et al. focuses on farmers' knowledge and perceptions of cocoa tree pests and the resulting damage, as well as the implications for pest management in cocoa plantations in Ghana. The results indicate that incorrect identification of major pests by over 85% of farmers has led to significant variation in pesticide application timing. The study underscores the importance of pest identification and monitoring as essential components of Integrated Pest Management (IPM) [12] Similarly,[13] evaluated farmers' adoption of recommended measures against cocoa swollen shoot virus disease (CSSVD). The results showed that farmer awareness significantly influenced the likelihood and extent of adopting these measures. However, adoption of recommended treatment procedures remained very low, highlighting the need for intensive farmer education by the Ghanaian government [13]. Alongside traditional methods, technological advancements play a crucial role in cocoa disease detection. Convolutional Neural Networks (CNNs) are powerful deep learning algorithms specifically designed for image processing, using deep learning to perform generative and descriptive tasks. In this regard, [14] used CNNs for the detection and classification of cocoa pod diseases. Their methodology includes building an image database, preprocessing images, and training CNN models. The models were evaluated on metrics such as mean Average Precision (mAP) and detection accuracy, demonstrating high performance in disease detection and classification [14]. Additionally, [15] employed a deep learning algorithm for the classification of cocoa pods into three states: healthy, infected with black pod disease, and attacked by a pest. The model achieved 94% accuracy but requires improvements for real-world applications [16] also tackled cocoa plant disease issues using CNNs. They evaluated five CNN architectures and found that the Custom CNN model performed best. However, they highlighted class imbalance in the dataset as a major limitation . To overcome CNN limitations, Capsule Networks (CapsNets) offer a promising alternative. CapsNets overcome CNN limitations by using capsule units that retain pose information and spatial relationships, providing a richer feature representation. [17] used CapsNets to address the problem of image super-resolution. Their SRCaps model converts low-resolution images into high-resolution ones by leveraging the unique capabilities of CapsNets. The results on the PlantVillage dataset were very satisfactory. Moreover, [2] explored an approach combining CapsNets with residual networks (ResNets) to enhance the classification capabilities of plant leaf diseases. The results showed that the CapsNet-ResNet model outperformed traditional CNN models. CapsNets have found numerous applications due to their general applicability in 2D deep learning. For example, [19] developed a deconvolutional capsule network, SegCaps, to tackle object segmentation . Furthermore,[20] extended CapsNets to segmentation and classification of actions by introducing capsule-pooling. Similarly, [21] , [22] , and, [23] . proposed Capsule-GANs, variants of capsule networks from standard Generative Adversarial Networks (GANs). These showed improved performance in 2D image generation . Finally, [24] demonstrated that capsule representations learn more meaningful 2D embeddings than neurons in a standard CNN. To our knowledge, the use of 3D data for detecting diseases in cocoa plants had not been explored before. To address this gap, our methodology proposes integrating 3D data to capture the three-dimensional structure of cocoa pods. Additionally, we enhance our approach by incorporating a self-attention mechanism into the routing process of CapsNets, aiming to improve detection accuracy by leveraging spatial information and local features.

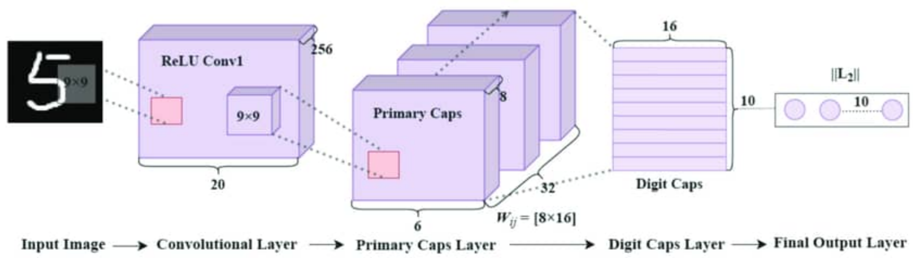

Fig. 1. Architecture des reseaux de neurones en capsule.

Originally proposed by [6] , the Capsule Network (CapsNet) addresses the issue of information loss regarding position due to the pooling layer in traditional CNNs. Traditional CNNs typically achieve the desired results when recognizing images that are similar to the training dataset. However, their recognition performance tends to decline when there is some degree of rotation, distortion, or changes in the relative positions of the images. Traditional CNNs focus on extracting local features and lack robustness to global changes in the image. Additionally, pooling operations can further blur positional information , compromising performance for images with significant spatial variations. Recent research [18] , [7] , [8] , [9] , [10] , [11] has extensively demonstrated the effectiveness of Capsule Networks (CapsNets) in various fields, particularly in image recognition and complex data analysis. Unlike Convolutional Neural Networks (CNNs), CapsNets fully integrate spatial relationships between objects and their components, thanks to their unique hierarchical architecture. Each layer of capsules in a CapsNet is designed to represent not only the presence of specific entities but also their attributes and relationships with other elements in the image. This hierarchical capability enables CapsNets to capture detailed information about the structure and orientation of objects, significantly enhancing the accuracy of classifications and recognitions. Today, their crucial importance in detecting plant diseases is undeniable. By enabling fine analysis of visual characteristics in leaves and fruits, CapsNets can early detect subtle signs of diseases such as spots, deformations, or growth anomalies, thereby assisting farmers in taking preventive measures and effectively improving overall crop health sustainably

Capsule Networks (CapsNets) stand out for their key concepts and innovative architecture. Fundamentally, a capsule represents a group of neurons designed to identify and characterize specific entities within input data. Unlike convolutional neural networks (CNNs), CapsNets integrate capsules that not only detect the presence of entities but also capture their spatial and relational attributes. Each layer of capsules in a CapsNet is organized hierarchically, where each lower-level capsule predicts activations of higher-level capsules through linear transformations. The mechanism of dynamic routing, known as “Routing by Agreement,” allows lower-level capsules to determine which higher-level capsules will receive their outputs, iteratively adjusting agreement coefficients to maximize consistency between predictions and outputs.

Mathematically, the output of a capsule $\mathbf{v}_j$ is computed by applying a “squashing” function to weighted activations from lower-level capsules:

$$\mathbf{v}_j = squash(\mathbf{s}_j) = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}$$

where $\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}$ is the weighted sum of predictions from lower-level capsules $\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i$, with $\mathbf{W}_{ij}$ representing the weight matrix between capsules.

Dynamic routing is achieved through iterative adjustment of agreement coefficients c_{ij} at each propagation step:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}$$

where b_{ij} starts at zero and is updated based on agreement between predictions $\hat{\mathbf{u}}_{j|i}$ and outputs $\mathbf{v}_j$:

$$b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i} \cdot \mathbf{v}_j$$

This complex mathematical approach enables CapsNets to effectively capture spatial and structural relationships among entities in data, offering a powerful representation capability for applications such as object recognition and, notably, precise detection of plant diseases.

2.1 proposed method

This part presents our contribution to cocoa pod disease detection using an innovative approach combining 3D images and Capsule Neural Networks (CapsNets) enhanced with self-attention mechanisms. We explain in detail the methodologies employed, the algorithms developed, and the results obtained. This work offers novel solutions to overcome limitations of existing 2D image-based methods by addressing subtle yet crucial features of cocoa pods such as swellings and minor deformations.

The contribution of this chapter is significant for several reasons. Firstly, it introduces a new approach to cocoa pod disease detection using 3D images that capture depth and structural details crucial for identifying anomalies. This overcomes limitations of traditional 2D methods that often fail to detect subtle yet important anomalies. Secondly, integrating self-attention in CapsNets significantly enhances detection accuracy by focusing model resources on relevant features. This technical advancement contributes to improving the state-of-the-art in plant disease detection, with potentially profound implications for agriculture and plant pathology research.

Methodology

To obtain 3D images of cocoa pods, we utilized a dataset of images downloaded from Kaggle. This dataset was created using advanced depth extraction techniques such as photogrammetry and 3D scanning, ensuring that the images fully incorporate necessary depth information for our study. Photogrammetry involves taking multiple 2D photos from different angles and stitching them together to create a 3D reconstruction. 3D scanning, on the other hand, uses devices like laser scanners or time-of-flight cameras to capture depth and spatial details of cocoa pods. These methods allow us to obtain precise and detailed 3D representations of cocoa pods, essential for subsequent processing and analysis steps.

The obtained 3D images are converted into point clouds, a representation where each point has 3D coordinates (x, y, z) describing the scanned object's surface. Preprocessing of point clouds includes several critical steps to enhance data quality and facilitate segmentation. We apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to segment these point clouds into density-based clusters. DBSCAN automatically identifies dense structures and separates regions of interest from background noise. This segmentation is crucial for isolating important features such as swellings and deformations indicative of diseases. The resulting clusters highlight regions of interest on cocoa pods, facilitating anomaly detection.

Modeling with Capsule Neural Networks (CapsNets)

Capsule Neural Networks (CapsNets) are designed to preserve hierarchical and spatial information of data. Unlike traditional CNNs, CapsNets use "capsules" that represent vector outputs instead of scalar neurons. In our project, we employ a 3D Capsule Network architecture to detect anomalies in 3D images of cocoa pods. Here's a detailed description of this architecture and its operation.

Before entering the Capsule Network, 3D images of cocoa pods are preprocessed to normalize pixel values and augment data, enhancing model robustness. These preprocessed images are then transformed into point clouds to better capture the 3D structure of pods. The DBSCAN algorithm is used to segment relevant clusters and filter out noise, allowing the model to focus attention on significant areas of 3D images.

The process starts with a 3D convolutional layer that extracts local features from segmented volumes. These features are encapsulated to form primary capsules. Each primary capsule detects simple features such as local contours or textures within segmented volumes and produces a pose vector. A pose vector is a vector representation that encodes both the presence and spatial configuration (position, orientation) of these local features. Its purpose is to capture rich information about the arrangement of features in the image, crucial for 3D object recognition.

After primary capsules, a squashing function is applied to normalize output vectors, scaling them between 0 and 1. This helps express the probability of detected features' presence while preserving their directions. The squashing function transforms outputs of primary capsules into vectors whose length represents the probability of certain features being present in the image. The mathematical formula of this function is:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}$$

where s_j is the weighted sum of inputs received by the capsule.

Next, a dynamic routing algorithm is employed to optimally route predictions from primary capsules to class capsules. This algorithm adjusts connection weights between capsules based on prediction probability and reliability, enhancing prediction accuracy and consistency. Integration of a self-attention mechanism allows weighting capsules based on their relative importance, enabling the model to focus on the most relevant areas of data.

Class capsules utilize aggregated information from primary capsules to form complex representations of anomalies in 3D images. Each class capsule produces a pose vector representing the probability of a specific anomaly being present in the image. These capsules integrate information from primary capsules to detect intricate entities.

To evaluate reconstruction quality and guide model learning, output capsules are used to reconstruct original 3D images. Fully connected layers with ReLU and Sigmoid activations are used to generate the reconstructed image. The Capsule Network's output is also used for anomaly classification. The total loss combines magnitude loss and reconstruction loss. Magnitude loss evaluates accuracy in anomaly classification, while reconstruction loss assesses quality of the reconstructed image.

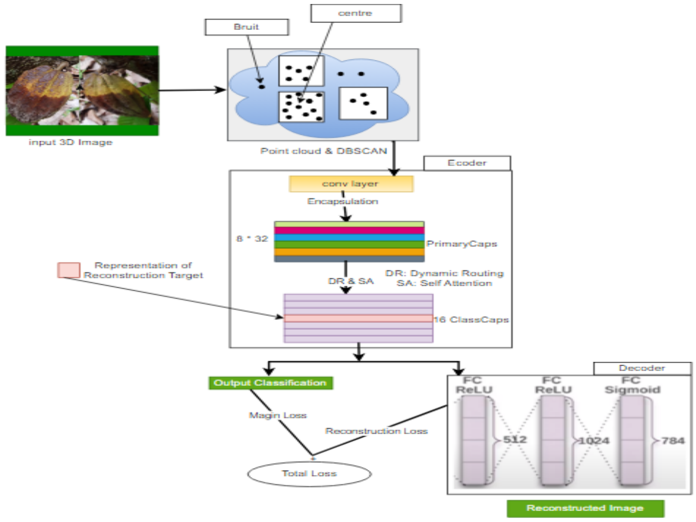

Fig. 2. 3D Architecture.

Margin loss is designed to measure the difference between network predictions and ground truths. It penalizes errors by adjusting network weights to minimize these differences, guiding the network to learn accurate detection of present and absent anomalies in images. The formula for margin loss is:

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2$$

where:

- T_k indicates presence (1) or absence (0) of class k anomaly in the image.
- $\|v_k\|$ is the length of output vector from class capsule for class k.
- m^+ is the upper threshold for desired vector length (e.g., 0.9).
- m^- is the lower threshold for desired vector length (e.g., 0.1).
- λ is a weighting coefficient (e.g., 0.5) balancing contributions of positive and negative terms in loss.

Reconstruction loss evaluates the quality of the image reconstructed by the capsule network. It compares the reconstructed image to the original image and penalizes the differences between the two. The reconstruction loss function is usually defined as the mean squared error (MSE) between the pixels of the original image and those of the reconstructed image:

$$L_{reconstruction} = \sum_i (x_i - \hat{x}_i)^2$$

where:

- x_i represents pixel i of the original image.

- $\hat{x}_i$ represents pixel i of the reconstructed image.

The total loss is a linear combination of margin loss and reconstruction loss. It is calculated as follows:

$$L_{total} = L_{margin} + \alpha \cdot L_{reconstruction}$$

where:

- L_{margin} is the margin loss.
- $L_{reconstruction}$ is the reconstruction loss.
- α is a weighting coefficient (e.g., 0.0005) balancing the contributions of the two loss terms.

By combining these two components, the total loss function trains the capsule network to improve both the accuracy of anomaly classification and the quality of 3D image reconstruction.

2.2 integration of the self - attention mechanism

To improve the performance of our model, we integrated a self-attention mechanism after the primary capsules. Self-attention allows the model to focus on the most relevant parts of the data by weighting the capsules according to their relative importance. Here is how we proceeded:

- **Calculation of Attention Scores**: We calculated the attention scores using the pose vectors produced by the primary capsules. For each capsule i, we used a weight matrix W_q to obtain a query (q_i), a weight matrix W_k to obtain a key (k_i), and a weight matrix W_v to obtain a value (v_i).

$$q_i = W_q \cdot pose_i, \quad k_i = W_k \cdot pose_i, \quad v_i = W_v \cdot pose_i \tag{1}$$

- **Calculation of Similarity Scores**: We calculated the similarity scores between the capsules using the dot product between the queries and keys, normalized by the square root of the key dimension d_k.

$$score_{ij} = \frac{q_i \cdot k_j}{\sqrt{d_k}} \tag{2}$$

- **Application of Softmax Mask**: We applied a softmax function to the similarity scores to obtain the attention weights.

$$\alpha_{ij} = softmax(score_{ij}) = \frac{\exp(score_{ij})}{\sum_j \exp(score_{ij})} \tag{3}$$

- **Weighting of Capsules**: The weighted values (v_i) were obtained by multiplying each value by its corresponding attention weight.

$$attention_i = \sum_j \alpha_{ij} v_j \tag{4}$$

– **Combination of Capsules**: The output vectors of the primary capsules were combined into a new representation using the values weighted by the attention scores.

$$capsule\_output_i = \sum_j \alpha_{ij} pose_j \tag{5}$$

– **Complete Formula**: The set of steps can be combined into a complete formula for the attention as follows:

$$capsule\_output_i = \sum_j \left(\frac{\exp\left(\frac{(W_q \cdot pose_i)\cdot(W_k \cdot pose_j)}{\sqrt{d_k}}\right)}{\sum_k \exp\left(\frac{(W_q \cdot pose_i)\cdot(W_k \cdot pose_k)}{\sqrt{d_k}}\right)} \right) (W_v \cdot pose_j) \tag{6}$$

3 Experimentation and Results

3.1 Implementation of Our Approach

Working Environment This section presents the tools and technologies (programming language, framework, interface, library) used to develop our model. In this section, we introduce the various tools we use for implementing our segmentation model. These tools include: the Python language, TensorFlow, NumPy libraries, etc. Python, an easy-to-learn programming language, offers high-level data structures and a vast standard library, making it ideal for rapid application development across various platforms [25]. Its concise syntax and dynamic typing simplify machine learning tasks, supported by extensive libraries (TensorFlow, Theano, Torch) and a large community.

TensorFlow, developed by Google Brain, is particularly suitable for high-performance computations, including image and voice recognition, video analysis, and text applications. It supports parallelism, reduces error rates by up to 60% in machine learning, and provides frequent updates.

The Keras library facilitates experimentation with deep neural networks via TensorFlow and Theano, focusing on usability and extensibility. NumPy, an essential library for numerical computation, enables efficient handling of high-performance multidimensional arrays.

For development, we use Jupyter Notebook, which supports Python, Julia, and R languages. Due to the segmentation model's reliance on convolutional neural networks, we require high-performance hardware to execute billions of operations. While CPUs handle tasks demanding high memory and connectivity, GPUs offer superior raw computing power, ideal for large-scale data analysis.

Implementation We began by generating a dataset of 3D point clouds, where each point cloud consists of 32x32x32 points representing cocoa pods. These points, initially randomly distributed, are then normalized to fit the specified dimensions and random noise is added to simulate realistic conditions, making the model more robust to variations in real data. The point clouds are segmented into clusters using the DBSCAN algorithm, which allows identifying clusters of arbitrary shapes and ignoring noise. To improve the model's performance and generalization, we apply data augmentation techniques, including rotation, resizing, reflection, and the addition of extra noise. Additionally, the point clouds are normalized, ensuring a uniform data distribution and accelerating convergence during training.

The model architecture begins with 3D convolutional layers to extract spatial features from the point clouds. These layers are followed by batch normalization and pooling layers to reduce dimensionality and control overfitting. We then integrate capsule networks to better model complex entities and their spatial relationships. An attention mechanism is added to highlight important parts of the point clouds, enhancing the model's ability to identify diseases. For training the model, we use a focal loss function to handle class imbalances in the data, and the Adam optimizer for its efficiency and ability to dynamically adjust the learning rate. The point clouds and cluster labels are separated into training and validation sets, then organized into batches for efficient processing by the model. Training is conducted over multiple epochs, with constant monitoring of performance on the validation set to avoid overfitting, and performance metrics are used to adjust hyperparameters and improve the model architecture as needed. This approach promises to provide more accurate and robust tools for detecting and managing diseases in cocoa plantations.

3.2 Experimental Results and Analysis

Performance Evaluation We used accuracy, F1 score, loss curves, and learning curves to evaluate our CapsNet model for detecting cocoa plant diseases. Accuracy provides a general measure of correct predictions, while the F1 score balances precision and recall, particularly useful for imbalanced classes. Loss curves visualize training and validation loss to identify issues like overfitting, and learning curves track performance over time to diagnose bias or variance. Combining these metrics ensures our model accurately detects true positives and minimizes false positives, while effectively monitoring its learning and overall performance.

Table 1. Performance of the Cocoa Pod Disease Detection Model

Metrics	Training	Test
Loss	0.1	Validation (0.6 to 0.1)
Accuracy	98%	95%
F1 Score	99%	98%
Rappel	99%	98%

First, we can see that the model achieves very good performance during the training phase, with an accuracy of 98%. This indicates that the model was able to learn and reliably identify the relevant features for detecting the various diseases affecting cocoa plants.

Regarding the model's generalization capability, the results on the test data are also very satisfactory. We observe an accuracy of 97%, which demonstrates that the model is capable of transferring its knowledge to new data. This suggests that the choices made in the model's design, such as the use of 3D inputs and the self-attention mechanism, were judicious and helped improve generalization capability.

The F1 score, which combines precision and recall, reaches 99% on the training set, confirming the model's excellent performance on these data. On the test data, the F1 score is 75%, which remains a solid performance despite a slight decrease compared to training.

Finally, the loss metric shows a value of 0.1 on the training set, indicating a very low error on these data. On the test set, this loss metric is maintained within the interval [0.6 to 0.1], confirming the model's good generalization.

Overall, these results are very encouraging and demonstrate the effectiveness of the developed model for detecting diseases affecting cocoa plants. The integration of the self-attention mechanism and the use of 3D inputs seem to have positively contributed to the model's performance, allowing it to acquire better generalization capability.

Curves and Interpretations The figures show that the deep learning approach applied to capsule neural networks is very promising for detecting certain diseases in cocoa plants. CapsNets enhanced by a self-attention mechanism demonstrated high accuracy in identifying diseases, surpassing traditional CNN models.

The use of 3D images as input allowed for considering the three-dimensional characteristics of cocoa pods, which improved the accuracy of disease detection. Additionally, the self-attention mechanism exploited both spatial information and local features, further enhancing detection accuracy.

These results suggest that this approach could be used to develop more accurate and efficient disease detection systems for cocoa plants, which could help improve cocoa production and reduce economic losses related to diseases.

- **Learning and Loss Curve**

The deep learning approach using Capsule Neural Networks (CapsNets) with a self-attention mechanism shows high promise for detecting cocoa plant diseases, surpassing traditional CNN models in accuracy. Utilizing 3D images has improved detection by considering the three-dimensional characteristics of cocoa pods, and the self-attention mechanism has further enhanced accuracy by exploiting spatial information and local features. These results indicate that this approach could lead to more accurate and efficient disease detection systems, improving cocoa production and reducing economic losses.

- **Learning and Loss Curve**

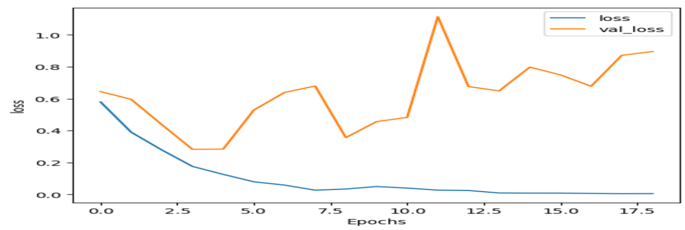

Fig. 3. Training and Validation Loss

From the curves, we can clearly see that at the beginning of the training, there is a high loss on the training set (train loss) while the loss on the validation set (val_loss) is lower. This indicates that the model is still struggling to accurately identify disease patterns in the data.

However, as the iterations progress, we observe that the loss decreases, and the validation loss, despite some fluctuations, generally follows a downward trend initially. The regular decrease in training loss indicates that the model is effectively learning from the training data. The fluctuations in the validation curve are common in the early training phases and may reflect the model's complexity in capturing the nuances of the validation data. Overall, this curve shows that our model is making progress in learning and could benefit from further refinement to stabilize performance on the validation set.

– **F1 Score Curve**

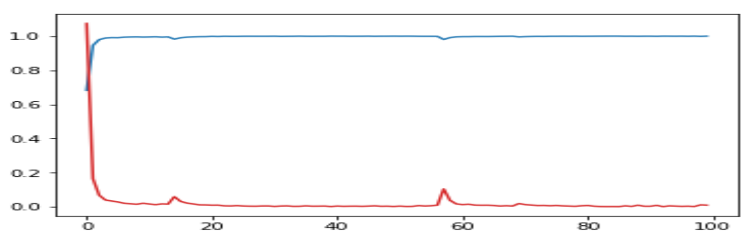

Fig. 4. F1 Score

From this graph, we can observe the performance of our capsule neural network model for detecting cocoa plant diseases in terms of F1 Score on the training data (orange curve) and the test data (blue curve). The blue curve shows an F1 Score that quickly reaches a high value, around 1 (or 99%), indicating excellent classification performance on the test data from the early iterations, and remains stable thereafter. The orange curve, representing the training data, drops rapidly to a value close to zero before climbing back up and stabilizing at a very low level, indicating that the model effectively learns from these data. The minor fluctuations observed towards the end of the graph show some variations, but overall, the F1 Scores demonstrate the model's robustness and reliability, achieving good performance (98%) on the test data and significant consistency on the training data. This confirms that our capsule neural network model is promising for the accurate detection of cocoa plant diseases.

- **Accuracy Curve**

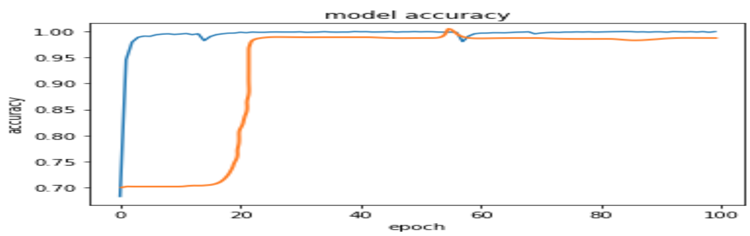

Fig. 5. accuracy on training and test data

From this graph, we can observe the performance of our capsule neural network model with an attention mechanism for cocoa plant disease detection in terms of accuracy on the training data (blue curve) and test data (orange curve). The blue curve shows accuracy rapidly reaching a high value close to 1 (99%), indicating excellent classification performance on the training data from the early iterations, and remaining stable thereafter. The orange curve, representing the test data, starts at around 70% and then gradually increases to stabilize around 95%. This trend demonstrates that our model effectively learns from the 3D training data, which includes depth information, and maintains very high performance on the test data. Overall, the accuracy performance reflects the robustness and reliability of our model, confirming that it is well-suited for precise cocoa plant disease detection, with accuracy rates approaching 100% on the training data and around 95% on the test data.

4 Comparison of our approach with the state-of-the-art

Three state-of-the-art articles [2], [17], [24] have applied capsule neural networks (CapsNets) on 2D images. The following table highlights the contributions of our method compared to these works.

Table 2. Comparison of Model Performance with State-of-the-Art Methods

Metrics	Test 3D et SA	Test 2D
loss	(0.6 à 0.1)	ND
precision	95%	98%
F1 score	98%	83.05%
Reminder	98%	83,05%

5 Conclusion

We developed an innovative approach for detecting cocoa pod diseases using Capsule Neural Networks (CapsNets) enhanced with a self-attention mechanism. Our work addressed the limitations of traditional methods by accurately detecting subtle anomalies, such as swellings or deformations in 3D images of cocoa pods. By combining advanced segmentation and modeling techniques, we optimized detection performance. The 3D images were converted into point clouds and segmented using the DBSCAN algorithm to identify important characteristics. We designed an enhanced CapsNet architecture with self-attention, allowing the model to dynamically focus on the most relevant areas. Experimental results demonstrated that our model significantly improved the accuracy of anomaly detection compared to traditional methods, offering a robust and precise solution for crop management and reducing economic and ecological losses.

References

1. Pugh-Moreno, T. *Etude des déséquilibres de liaison dans une collection de cacaoyers (Theobroma cacao L.) appartenant au groupe Criollo/Trinitario et application au marquage génétique des caractères d'intérêt.* Doctoral dissertation, ENSAM, 2005.
2. Zhang, X., Mao, Y., Yang, Q., Zhang, X. (2024). A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network. IEEE Access, 12, 44573–44585. doi:10.1109/ACCESS.2024.3377230.
3. Sabour, S., Frosst, N., Hinton, G. E. *Dynamic Routing Between Capsules.* NIPS, 2017.
4. Xi, E., Bing, S., Jin, Y. *Capsule Network Performance on Complex Data.* arXiv preprint arXiv:1712.03480, 2017.
5. Afshar, P., Mohammadi, A., Plataniotis, K. N. *Brain Tumor Type Classification via Capsule Networks.* IEEE Access, 2018.
6. Hinton, G. E., Krizhevsky, A., Wang, S. D. *Transforming Auto-encoders.* ICANN, 2011.
7. G. Altan, "Performance evaluation of capsule networks for classification of plant leaf diseases," *International Journal of Applied Mathematics, Electronics and Computers*, vol. 8, no. 3, pp. 57–63, Oct. 2020.
8. V. H. Kondekar and S. K. Bodhe, "Automation in plant pathology: Optimized attentional Capsule_BiLSTM optimized with chaotic sparrow algorithm for colour feature-based plant disease detection," *Multimedia Tools and Applications*, early access, Oct. 2023.

9. L. M. Abouelmagd, M. Y. Shams, H. S. Marie, and A. E. Hassanien, "An optimized capsule neural networks for tomato leaf disease classification," *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, p. 2, Jan. 2024.
10. Y. Li, M. Qian, P. Liu, Q. Cai, X. Li, J. Guo, H. Yan, F. Yu, K. Yuan, J. Yu, L. Qin, H. Liu, W. Wu, P. Xiao, and Z. Zhou, "The recognition of Rice images by UAV based on capsule network," *Cluster Computing*, vol. 22, no. S4, pp. 9515–9524, Jul. 2019.
11. S. Sapna, S. Sandhya, V. Acharya, and V. Ravi, "Apple foliar leaf disease detection through improved capsule neural network architecture," *Multimedia Tools and Applications*, early access, Nov. 2023.
12. Kweku Awudzi, G., Adu-Acheampong, R., Avicor, S., Bukari, Y., Yeboah, M., Boateng, E., Ahadzi, S. (2021). Farmers' knowledge and perception of cocoa insect pests and damage and the implications for pest management on cocoa in Ghana. Journal of Plant Protection Research, 61, 145-155. doi:10.24425/jppr.2021.137022.
13. Amon-Armah, F., Domfeh, O., Baah, F., Owusu-Ansah, F. (2021). Farmers' adoption of preventive and treatment measures of cocoa swollen shoot virus disease in Ghana. Journal of Agriculture and Food Research, 3, 100112.
14. 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:10.1016/j.procs.2022.07.013.
15. Godmalin, R., Aliac, C., Feliscuzo, L. (2022). Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm. In Proceedings of the 2022 International Conference on Artificial Intelligence and Emerging Technologies (IICAIET), 1-5. doi:10.1109/IICAIET55139.2022.9936817.
16. 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, 7732–7741.
17. Araújo, G. C. de, Jordão, A., Pedrini, H. (2023). Single Image Super-Resolution Based on Capsule Neural Networks. In Proceedings of the Brazilian Conference on Intelligent Systems, 113–128. Springer Nature Switzerland, Cham.
18. S. Verma, A. Chug, and A. P. Singh, "Exploring capsule networks for disease classification in plants," *Journal of Statistics and Management Systems*, vol. 23, no. 2, pp. 307–315, Feb. 2020.
19. LaLonde, R., Bagci, U. (2018). Capsules for Object Segmentation. arXiv preprint arXiv:1804.04241.
20. Kevin Duarte, Yogesh Rawat, Mubarak Shah. *VideoCapsuleNet: A Simplified Network for Action Detection.* In *Advances in Neural Information Processing Systems*, pages 7621–7630, 2018.
21. Ayush Jaiswal, Wael AbdAlmageed, Yue Wu, Premkumar Natarajan. *CapsuleGAN: Generative Adversarial Capsule Network.* In *European Conference on Computer Vision*, pages 526–535, Springer, 2018.
22. Raeid Saqur, Sal Vivona. *CapsGAN: Using Dynamic Routing for Generative Adversarial Networks. arXiv preprint arXiv:1806.03968*, 2018.
23. Yash Upadhyay, Paul Schrater. *Generative Adversarial Network Architectures for Image Synthesis Using Capsule Networks. arXiv preprint arXiv:1806.03796*, 2018.
24. Ancheng Lin, Jun Li, Zhenyuan Ma. *On Learning and Learned Representation with Dynamic Routing in Capsule Networks. arXiv preprint arXiv:1810.04041*, 2018.
25. Ting, Kai Ming and Witten, Ian H, *Issues in stacked generalization*, Journal of Artificial Intelligence Research, 1999, vol. 10, pp. 271-289.