International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 1-12
https://doi.org/10.70917/ijcisim-2026-2658
Article

Dragonfruit Stem Health Classification with Deep Learing and Attention Mechanisms

Ashwini Prasad S[1], Uma S[2], Sheba Pari N[3]*

[1]School of Computer Science and Engineering, RV University, Vidyaniketan Post, Bangalore, Karnataka, India. Email: ashwinips@rvu.edu.in
[2]Department of Computer Applications, B.M.S. College of Engineering (Affiliated to Visvesvaraya Technological University), Bangalore, Karnataka, India. Email: uma.mca@bmsce.ac.in
[3]School of Computer Science and Engineering, RV University, Vidyaniketan Post, Bangalore, Karnataka, India. Email: shebapari@rvu.edu.in
***Correspondence author:** Ashwini Prasad S

Abstract: Plant disease detection is important for maintaining the health and quality of crop yields. For the detection of health problems in images of dragon-fruit (Hylocereus) stems, we present a deep learning architecture that is aug-mented by the application of spatial, channel, and domain-specific attention mechanisms. For the purpose of improving accuracy and robustness, the model architecture is combined using features from ResNet50 and EfficientNetB0 back-bones, along with separate attention branches. Training and testing were per-formed using a proprietary dataset of images of dragonfruit stems that are healthy as well as showing various levels of disease. The model initially used frozen fea-ture extractors for training, which were subsequently fine-tuned to improve over-all performance. Experimental results provide high accuracy for classification, with ROC-AUC values above 0.94 for all classes. The proposed method enables precision agriculture operations by providing a robust method for early detection of dragonfruit diseases. Future work will include adapting the model to real-world field images and further optimising it for application in agricultural moni-toring systems.

Keywords: Plant disease classification, hybrid deep learning, ResNet50, Effi-cientNetB0, attention mechanism, Grad-CAM, interpretability, classification accuracy, early disease detection, robustness

1. INTRODUCTION

Early detection of plant diseases is crucial to the achievement of agricultural produc-tivity in precision agriculture, an accomplishment made easier by recent technology breakthroughs in the agricultural sector. Tropical dragonfruit (Hylocereus spp.), one of the increasingly valuable fruits around the world in terms of its commercial value, has shown immense potential based on its nutritional benefits, distinctive outlook, and adaptability to a wide range of climatic conditions. However, just like with the majority of other horticultural produce, dragonfruit production is extremely vulnerable to a myr-iad of stem-based diseases with significant impacts on the quality as well as quantity of yields. These stem diseases are prone to show themselves subtly in the initial stages, thereby making it imperative to conduct lengthy and human-error-prone visual checks by specialists, particularly when used to assess large-scale plantations. There is thus an urgent need for advanced plant health monitoring systems, and as such, the glaring need for automated, efficient, and precise disease detection methods in dragonfruit produc-tion.

Conventional plant disease detection techniques have mostly depended on subjective laboratory tests and manual visual inspection by hand, both of which are non-scalable in commercial use. The ability to identify and classify plant diseases from visual char-acteristics in images has greatly evolved with recent advancements in computer vision methods, specifically deep learning algorithms. Studies with the use of convolutional neural networks (CNNs) in the detection of different plant diseases have been promis-ing [5][7]. However, few studies have been performed on the stems of dragonfruit, which present distinctive visual manifestations of disease progression different from those reported in leaves or fruits. With the differences in texture variations, colour gra-dients, and patterns of lesion

distributions in stems—factors commonly overlooked when leaf images dominate the dataset—prior studies such as [6] and [7] have docu-mented that disease identification in stems involves specialised techniques.

Current deep learning-based models still face many challenges to be addressed despite remarkable advancements. They include overfitting due to short or unbalanced da-tasets, poor generalisation to actual field conditions, and absence of model interpreta-bility, which restricts their practical application [1][3]. For instance, the research of Amara et al. [1] highlights how important explainability is for plant disease classifiers to be capable of winning farmers' confidence and enhancing acceptance in real-world applications. In addition, most existing solutions utilise traditional CNNs, unable to adequately address the intricate spatial and channel-wise feature distributions typical of stem architectures, particularly in crops such as dragonfruit. While new architectures that utilise multiple attention mechanisms have promising advancements in other areas [2][4], nothing is known about how they could be applied to dragonfruit stem health monitoring specifically.

The low diversity in datasets, the overfitting tendency of the model, and the inability to learn from real field conditions, as emphasised in sources [1], [3], and [8], are the main shortcomings of the existing body of literature that this research specifically seeks to overcome. This research presents a novel hybrid attention-based deep learning architecture specifically designed for the identification of dragonfruit stem diseases, which is informed by lessons from previous research that emphasised model interpretability and the need for robust architectures capable of dealing with variability in agricultural data [1][6][9]. The employed model entails a multi-branch attention architecture that operates independently to obtain spatial, channel, and personalised feature representa-tions, thereby deviating from the traditional CNN approaches that mainly focus on

linear feature extraction. To effectively classify the health of dragonfruit stems, the model is designed to ignore irrelevant background oscillations while emphasising sub-tle disease-related signs. This is achieved through the utilisation of these specialised attention modules, as feature extractors using ResNet50 and EfficientNetB0.

This study stands out for its methodology to boost the generalizability of models in spite of the inconsistencies in real-world dragonfruit datasets. The dataset presented considerable challenges because of class imbalance and intrinsic variability, even though it was designed to include a wide range of images representing healthy and diseased stems under different lighting conditions, plant growth stages, and levels of disease severity. To stabilise the learning of low-level features, initially, the model was trained with base networks frozen; then all layers were unfrozen and fine-tuned with a lower learning rate to adjust to the sophisticated distributions of high-level features. This two-stage training approach was used to meet these challenges. The fine-tuning method was used to provide stable performance even when faced with unseen stem samples with irregular visual characteristics, which is in accordance with methodolo-gies advocated in recent plant disease detection literature [14][15][19].

By the deployment of this holistic methodology, the research offers a field-deployable system for monitoring plant health tailored to dragonfruit that is always effective in a variety of cultivation environments and further improves detection accuracy. This so-lution overcomes one of the biggest limitations of current agricultural AI technologies since it not only offers the first generalised framework for evaluating the health of drag-onfruit stems but also pushes the technological frontiers of plant disease detection by employing hybrid attention mechanisms [2][6][9]. To further push precision agriculture efforts in dragonfruit cultivation, future research will investigate the scalability of the system to large-scale field deployment environments and the integration of modules for continuous learning and real-time image acquisition.

2. Literature Review

The incorporation of deep learning approaches into plant disease detection has made monitoring systems in agriculture far more accurate and scalable. Numerous studies have examined several aspects of automatic disease classification of leaves and fruits, highlighting strengths of rich datasets and resilient model architectures. Amara et al.

[1] highlighted the need for explainable deep learning models in agriculture by high-lighting interpretability through concept identification. Likewise, Benfenati et al. [2] introduced the strengths of multispectral imaging for early-stage disease detection by introducing unsupervised deep learning approaches for disease detection. Meanwhile, Gohil et al. [3] introduced a hybrid system that combined classification and localisation, highlighting real-time capabilities that are fundamental to field-level practicality. Hasin et al. [4] took this debate a notch higher by employing region-based convolutional neu-ral networks, highlighting the need for accurate object-level disease localisation. All of

these studies have provided a starting point for model design that can operate in uncon-trolled environments; however, they have only looked at leaf imaging and not vital areas like stems.

Although the importance of stem-based disease detection for crops such as dragonfruit has grown, research on this topic has been limited. Kumar and Singh [5] used traditional deep learning methods in their research on leaf image analysis but noted that the out-comes were lacking in terms of tackling non-uniform visual patterns, more common in stems. As a response to this limitation, Li et al. [6] and Liu et al. [7] introduced deep convolutional models for imaging plant stems, showing that the focus on spatial textural features resulted in better detection rates. Mahlein et al. [8] also gave a more general overview of the integration of robotics and optical sensors into disease management systems and posited that focusing more on the different components of plants would be a major factor in the determination of precision agriculture results. In their investigation of classification recognition issues, Meng et al. [9] focused on distinguishing between mature and newly emerging disease classes. This issue is particularly important when it comes to crops like dragonfruit, where stem visual signs dramatically change at dif-ferent stages of ripeness.

Recent research has been increasingly concentrating on dataset-related methodologies and data management. Mignoni et al. [10] and Moupojou et al. [11] introduced domain-specific datasets for the identification of pests and plant diseases, indicating the im-portance of domain-specific initiatives concerning data gathering. Research on the ap-plication of IoT and machine learning for disease monitoring was investigated by Ou-hami et al. [12], thereby creating opportunities for the development of in-field, real-time disease detection systems for dragonfruit. Contribution to the area was provided by Rajbongshi et al. [13], who designed a comprehensive dataset for guava, which is similar to dragonfruit in terms of patterns of disease manifestation in structural tissues. Research by Salman et al. and Saleem et al. [14], [15] investigated AI-based patholog-ical trends and performance-optimised models, providing valuable insights for im-proved robustness and generalisation in field-level practical applications. Sethy et al.

[16] exhibited the efficacy of feature-based support vector machines in rice disease classification, emphasising feature attention even in non-neural network structures.

Finally, improvements in methodology and larger reviews have further pushed the boundaries of the field. A systematic review of deep learning methods in plant pathol-ogy was performed by Shoaib et al. [17], who noted the trend towards multi-modal and hybrid systems. Singh et al. [18] presented the PlantDoc dataset, which enabled visual disease detection across a variety of crops to be enhanced. Xu et al. [20] showed YOLO-based models could improve detection speed and accuracy for deployment on a large scale, and Wang et al. [19] surveyed the newest deep learning applications in plant disease and pest detection, noting the need for model generalizability. These contribu-tions notwithstanding, little development of hybrid attention-based models exists for the analysis of stem images in crops like dragonfruit, where stem health directly corre-sponds to plant viability as a whole. This paper aims to fill the gap by presenting a new

hybrid attention-infused deep learning model for dragonfruit stem disease classifica-tion, ensuring accuracy and field-level deployment.

3. Methodology

Dataset

The data used in the current study is from the repository named "**Dragon Fruit Stem Disease: An Annotated High-Resolution Image Dataset for Classification and Seg-mentation**" and is publicly accessible on Mendeley Data. The data classifies dragon-fruit (Hylocereus spp.) stem images categorically into six main categories: Anthrac-nose, Brown Stem Spot, Grey Blight, Healthy, Soft Rot, and Stem Canker. Each cate-gory consists of colour images with varying disease severity and has been taken in var-ying real-world conditions, typified by varying lighting conditions, maturity of plants, and intricate backgrounds. The images also have natural ambient noise factors, includ-ing partial occlusions, non-uniform lighting, and surface debris, thus giving a realistic view of the field conditions typically faced by dragonfruit growers.

The collection is composed entirely of stem photos, a drastic departure from the typical plant disease collections, which are biased towards leaf photos. For dragonfruit, where stem condition is inextricably tied to plant health and yield potential, this emphasis on stems is especially important. The collection includes a considerable range of image sizes; some are taken in high definition, and others have unevenly framed and low-resolution images. This range of image size and quality not only reflects the actual dif-ficulties of field photography but also adds considerably to the difficulty of the prepro-cessing and model-training stages. The collection presents a strict and representative sample

set required for the evaluation and development of general disease detection algorithms especially designed for dragonfruit cultivation, as it includes a range of dis-ease appearances and natural environmental conditions.

Dataset Processing

The initial dataset of dragonfruit stems went through a long processing procedure that included dataset balancing, image normalisation, dimension change, and data augmen-tation to guarantee the fairness and accuracy of the training model. The initial dataset suffered from class imbalance, which is defined by unequal numbers of images in the six different classes. To ensure uniformity, each class ($c \in \{1,2, \dots ,6\}$) was limited to the number of images equal to the smallest sample size in all classes. Mathematically, the effective sample size per class, ($N_{\text{effective}}$), was defined as follows, where(N_c) is the

number of images in class (c).

$$N_{\text{effective}} = \min_{c\in\{1,2,6\}} N_c$$

The data were then divided into three distinct subsets—training, validation, and test-ing—after the class balancing process. To allocate 70% of the images to the training subset, 15% to the validation subset, and the remaining images to the testing subset, the sets were divided proportionally. Let the number of images in each class after the bal-ancing process be represented by (N_{total}). Then the allocation was calculated as:

$$N_{\text{train}} = \lfloor 0.7 \times N_{\text{total}} \rfloor, \qquad N_{\text{val}} = \lfloor 0.15 \times N_{\text{total}} \rfloor, \qquad N_{\text{test}} = N_{\text{total}} - N_{\text{train}} - N_{\text{val}}$$

where the floor operation to obtain integer sample sizes is represented by ($\lfloor \cdot \rfloor$). This splitting technique prevents class imbalance during training and testing from inducing bias and keeps the statistical integrity of the dataset intact.

To provide scale uniformity and data type suitable for deep neural network processing, encoding of images was used because of the variations in resolutions and formats of the original dragonfruit stem images. Normalisation of every image($I: \Omega \rightarrow R^3$) to a float-ing-point range ($[0,1]$) was performed, where (Ω) is the originally composed spatial domain of pixel intensities ($p(x, y, c) \in [0,255]$) per channel ($c \in \{R, G, B\}$)

Inormalized(x,y,c)=255p(x,y,c)

where

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{\Omega}, \forall \boldsymbol{c} \in \{\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B}\}$$

For improving numerical stability in gradient-based optimization in deep learning models, this normalization keeps input magnitudes within constant bounds.

A uniform size of (128×128) pixels was achieved by symmetric zero-padding after each image was reduced to its intrinsic aspect ratio. This was to keep the variance of the image size as small as possible. Denoting the original size of an image by ((h, w)), we calculated the scaling factor (s) for resizing as follows

s = min(128/h, 128/w)

The new dimensions ((h', w')) after resizing were therefore given by:

$h' = s \cdot h, \; w' = s \cdot w$

In accordance with achieving the fixed input size of the model and maintaining the aspect of the dominant visual components for dragonfruit stem classification purposes, the images were padded equally from both sides after scaling.

Data augmentation techniques were used to train the model to learn more and increase the diversity of the training data. The input images were randomly transformed with rotation, zoom, brightness, and flipping. In particular, a rotation matrix denoted as ($R\theta$)) was used to rotate an image according to pixel coordinates ((x, y)), and rota-tion (θ) was varied between $(-15^\circ) and (15^\circ)$.

$x' = x \cos\theta - y \sin$

$y' = x \sin\theta + y \; theta \cos$

Brightness changes included applying an arbitrary intensity change. ($\Delta_b \in [-0.2,0.2]$)

to change the last pixel brightness ($p'(x,y,c)$)according to the following formula:

$p'(x,y,c) = \text{clip}(p(x,y,c) + \Delta_b, 0,1)$

Zoom operations were randomly chosen within a uniform range.([0.9,1.1]). where pixel values are preserved in the range ([0,1]) due to (clip).

The model in question would be capable of learning major and distinctive features needed for effective health classification due to a thorough processing method for the dataset. The method included balancing, encoding, resizing, and augmentation. The method made it possible to ensure that the training would be capable of handling the enormous variations found in dragonfruit stem images taken in the field.

Attention Mechanisms Implementation

The proposed architecture employs three distinct attention mechanisms to enhance the image feature extraction from dragonfruit stems. The different aspects of visual infor-mation, which are critical to detecting disease, are captured by each attention branch, including spatial context, channel-wise importance, and local interaction of features. The feature ($A(x)$)augmented by attention derived from the input image (x) can be written as:

$$A(x) = \gamma\left(\mathcal{S}(f_s(x)) \odot f_s(x)\right) + \gamma\left(\mathcal{C}(f_c(x)) \odot f_c(x)\right) + \gamma\left(\mathcal{M}(f_m(x)) \odot f_m(x)\right)$$

Here, ($f_s(x)$), ($f_c(x)$), $and(f_m(x))$ are the spatial attention-based feature maps, chan-nel attention-based feature maps, and multi-head attention-based feature maps, respec-tively. Further, ($\mathcal{S}(\cdot)$), ($\mathcal{C}(\cdot)$), and($\mathcal{M}(\cdot)$)are spatial, channel, and multi-head attention-related operations, respectively. The notation ($\odot$) is used to represent element-wise multiplication. Further, ($\gamma(\cdot)$)represents a regularisation and normalisation operation that is used to attend to the features before aggregation.

The spatially pertinent areas of the input feature map, i.e., discolouration, lesions, and textural abnormality on the dragonfruit stem, are highlighted by the spatial attention mechanism. ($\mathcal{S}(\cdot)$). To create spatial maps along channel dimensions, both global av-erage pooling and global max pooling methods are utilised. Convolutional processes are then utilised to generate an attention map highlighting salient regions. The first spa-tial feature representation ($f_s(x)$)is utilised as an element-wise operation on the output attention map, thereby highlighting regions considered most significant in order to clas-sify the presence and intensity of the disease correctly. For the dragonfruit stem, where signs of sickness tend to appear in localised patchy patterns hard to describe with stand-ard convolutional layers alone, this spatial emphasis is especially crucial.

In contrast, along the channel axis of feature maps, the channel attention mechanism ($\mathcal{C}(\cdot)$) highlights or suppresses whole feature channels based on global significance. Shared multilayer perceptron maps channel-wise descriptors to low-dimensional space and then reconstructs them to accommodate the original channel size. Channel-wise descriptors are created by utilising both average pooling and max-pooling operations along spatial dimensions. The model can potentially select and focus on important col-our and texture representations inherent in various illness classes due to the weighted feature map.($\mathcal{C}(f_c(x)) \odot f_c(x)$). By highlighting subtle colour differences impercep-tible otherwise, channel attention greatly enhances the model's discriminative power for dragonfruit stems, where chromatic variation can represent different infection types.

By encoding relational relationships between different spatial elements of the image, the company-owned multi-head attention mechanism ($\mathcal{M}(\cdot)$) surpasses regular convo-lutional processes. This approach, inspired by self-attention architectures, employs the input feature map. ($f_m(x)$)to generate query, key, and value representations. Scaled dot-product operations are employed to compute attention weights, and weighted ag-gregation is employed to enrich feature representations. When applied during early-stage stem diseases, where infection signs are not necessarily visually apparent, the network can capture intricate intra-image interactions by employing numerous heads to attend to different sub-regions at the same time. A full, very informative attention-aug-mented feature representation($A(x)$) that is supplied to the model's subsequent stages is achieved by normalising and regularising the output of each attention mechanism via($\gamma(\cdot)$) to ensure numerical stability and consistency before aggregation.

Implementation Architecture

Stage	Operation	Output Dimension	Purpose
Input Layer	Input Image	128 × 128 × 3	Accepts resized and normalised stem images.
Branch 1: Spatial Attention	Convolution → BatchNorm → ReLU → Spatial Attention → MaxPooling → 1×1 Convolution	64 × 64 × 64	Extracts spatial dependencies in lesion regions.
Branch 2: Channel Attention	Convolution → BatchNorm → ReLU → Channel Attention → MaxPooling → 1×1 Convolution	64 × 64 × 64	Focuses on important spectral and textural features.
Branch 3: Custom Attention	Convolution → BatchNorm → ReLU → Custom Multi-Head Attention → MaxPooling → 1×1 Convolution	64 × 64 × 64	Captures complex relational patterns across image patches.
Feature Aggregation	Concatenation across the three branches	64 × 64 × 192	Merges multi-attention outputs.
Transition Layer	1×1 Convolution (Channel Reduction)	64 × 64 × 3	Matches the input format required by deep backbone networks.
ResNet50 Backbone	Pretrained ResNet50 Feature Extraction	2 × 2 × 2048 (then pooled)	Captures deep semantic features.
EfficientNetB0 Backbone	Pretrained EfficientNetB0 Feature Extraction	4 × 4 × 1280 (then pooled)	Extracts compact and efficient features.
Global Feature Fusion	Concatenation of ResNet50 and EfficientNetB0 outputs	1 × 1 × 3328	Combines rich semantic embeddings from both backbones.
Post-Processing Block	BatchNorm → Dropout (0.3)	3328	Regularises and stabilises the feature space.
Output Layer	Dense Layer → Softmax Activation	6 Classes	Predicts the dragonfruit stem health status.

The suggested dragonfruit stem health detection model adopts a hybrid multi-branch deep learning architecture that combines powerful backbone networks with specially designed attention mechanisms to facilitate effective feature extraction. The architec-ture consists of three parallel branches—spatial attention, channel attention, and spe-cially designed attention mechanisms—each of which is listed in Table 1. Subse-quently, the attention-refined features are merged and fed into ResNet50 and Efficient-NetB0, two deep feature extractors. Global pooling, batch normalisation, and regulari-sation are conducted on the merged outputs of the backbone networks before the final classification probabilities are generated using a softmax output layer.

The overall functional representation of the model can be expressed compactly as:

$$\hat{y} = \text{Softmax}\left(\phi\left(\mathcal{F}_{\text{ResNet50}}(g(\mathcal{A}_{\text{spatial}}(x), \mathcal{A}_{\text{channel}}(x), \mathcal{A}_{\text{custom}}(x))) \parallel \mathcal{F}_{\text{EfficientNet}}(g(\mathcal{A}_{\text{spatial}}(x), \mathcal{A}_{\text{channel}}(x), \mathcal{A}_{\text{custom}}(x)))\right)\right)$$

where:

(x) denotes the input dragonfruit stem image,

($A_{spatial}(\cdot)$), ($A_{channel}(\cdot)$), and ($A_{custom}(\cdot)$)represent the spatial, channel, and custom attention operations respectively,

($g(\cdot)$) denotes the concatenation operation across the attention-processed branches,

($\mathcal{F}_{\text{Rse50}}(\cdot)$)$and$($\mathcal{F}_{\text{Efcete}}(\cdot)$)denote feature extraction through ResNet50 and EfficientNetB0 respectively,

(‖) denotes feature concatenation,

($\boldsymbol{\phi}(\cdot)$) denotes the post-processing block comprising pooling, batch normali-sation, and dropout,

($\hat{y}$)is the predicted output probability vector for classification.

The central idea of this design is to allow the model to automatically pay attention to various visual features that are essential in the detection of diseases in dragonfruit stems. The three attention modules are designed to highlight complementary outputs: the tailored attention module highlights complex, localised patterns that are difficult to identify in the early stages of disease progression, the channel attention module high-lights salient colour and texture features, and the spatial attention module identifies location-specific discolouration and lesion patterns. A feature aggregation function ($g(\cdot)$) is used to combine the outputs from various attention modules such that the resulting feature map maintains the complex, multi-faceted representations required for successful learning.

After generating the integrated attention-refined features, these are simultaneously pro-cessed by ResNet50 and EfficientNetB0, two of the most popular deep feature extrac-tors renowned for their exceptional strengths in feature abstraction. In the post-

processing stage symbolised by ($\boldsymbol{\phi}(\cdot)$), the respective extracted representations are batch-normalised, concatenated, and dropped out for regularisation. This suppresses overfitting, enhances training stability, and enables better generalisation. For the final classification, a softmax activation function is applied to produce the probability distri-bution. ($\hat{y}$) over the six various classes corresponding to dragonfruit stem health. With the holistic nature of the architecture, the system becomes resilient to varying field con-ditions and disease severities, with localised patterns and overall semantic structures inherent in dragonfruit stem diseases being well represented.\

Training Strategy and Fine-tuning

The primary goals of the training method used for the suggested dragonfruit stem health diagnosis model were to attain high classification performance and exhibit greater gen-eralisation capability. To speed up training efficiency and avoid overfitting risk, a prev-alent issue in crop image data, a two-step training procedure was adopted [14][15][19]. Under the initial step, the attention branch and classification branch were allowed to learn task-related patterns while freezing the ResNet50 and EfficientNetB0 backbone networks in a way that their pretrained features can be exploited. With the first training step, only the attention modules, the fusion layers, and the output classifier head were trained while the parameters of the pretrained backbones were fixed.

Let (θ_{backbone}) denote the parameters of the backbone networks and let ($\theta_{\backslash new}$) denote the parameters of the newly initialised attention and classification layers. The optimi-sation process had the objective restricted to the following in the first phase of training:

$$\min \mathcal{L}\left(\theta_{\text{new}} \mid \theta_{\text{backbone rozen}}\right)$$

θ_{new}

The cross-entropy loss, calculated between the predicted and actual class distributions, is denoted as ($\mathcal{L}$). This enabled the creation of strong high-level representations that were in line with the nature of dragonfruit stem disorders, without compromising the effective low-level feature extraction capabilities inherited from large pretraining sets like ImageNet.

Following the convergence attained in the initial phase, the model underwent a fine-tuning phase where all the network parameters, including backbones, were left un-touched and optimised together at a considerably decreased learning rate. Fine-tuning enabled the entire network to keep specialising and fine-tuning for the task of classify-ing dragonfruit stem health. The following expression denotes the goal of the fine-tun-ing process:

$$\min \mathcal{L}\left(\theta_{\text{all}}\right) \quad \text{where} \quad \theta_{\text{all}} = \theta_{\text{backbone}} \cup \theta_{\text{new}}$$

θ_{all}

and where the learning rate ($\eta_{\text{fine-tune}}$) was set to:

$$\eta_{\text{fine-tune}} = \alpha \times \eta_{\text{initial}}, \qquad \alpha \in (0,0.1)$$

To enable stable convergence, the empirical value of (α) employed in this research was set at(10^{-2}), which is a reduction to one-hundredth of the initial learning rate. By fine-tuning both low-level and high-level features, this fine-tuning enhanced the per-formance of the model in identifying subtle structural and chromatic changes in drag-onfruit stems indicative of early-stage infections.

Several regularisation methods were used to further improve the training dynamics. To avoid overfitting, early stopping was used, which was based on tracking validation loss. As soon as performance levelled off on the validation set, a learning rate scheduler was used to dynamically reduce the learning rate. After the feature fusion step, dropout reg-ularisation with a rate of 0.3 was used to enhance generalisation by avoiding co-adap-tation of neurons. The addition of this training regimen allowed for robust predictive performance on a variety of difficult and diverse sets of dragonfruit stem disease through the preservation of balance between stability at the frozen stage and adaptabil-ity at fine-tuning.

4. Result Analysis

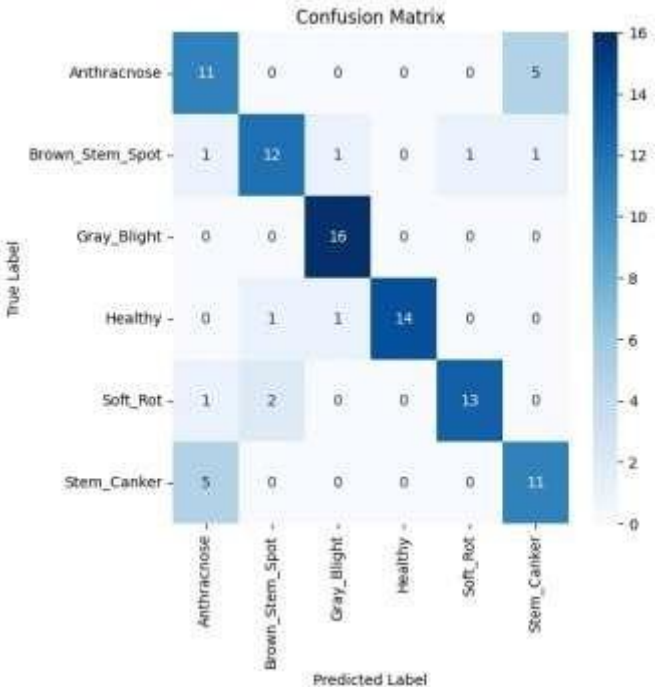

Figure 1 Confusion Matrix

The confusion matrix for the classification results on the dragonfruit stem test set is shown in Figure 1. All six classes had very good true positive rates, which indicate the model's very good overall performance. Notably, very little misclassification was seen, with the Grey Blight and Healthy classes being perfectly or nearly perfectly classified. There was some confusion between the Anthracnose and Stem Canker classes, espe-cially since cases of Stem Canker were more often confused with Anthracnose. This is to be expected, as the visual appearance of the onset symptoms of these two diseases may be similar and may include dark spots with irregular edges. Nonetheless, the pre-dominance of accurate classifications is very positive, which indicates that the proposed attention-augmented hybrid model can discern between the subtle signs of stem disease in dragonfruit crops.

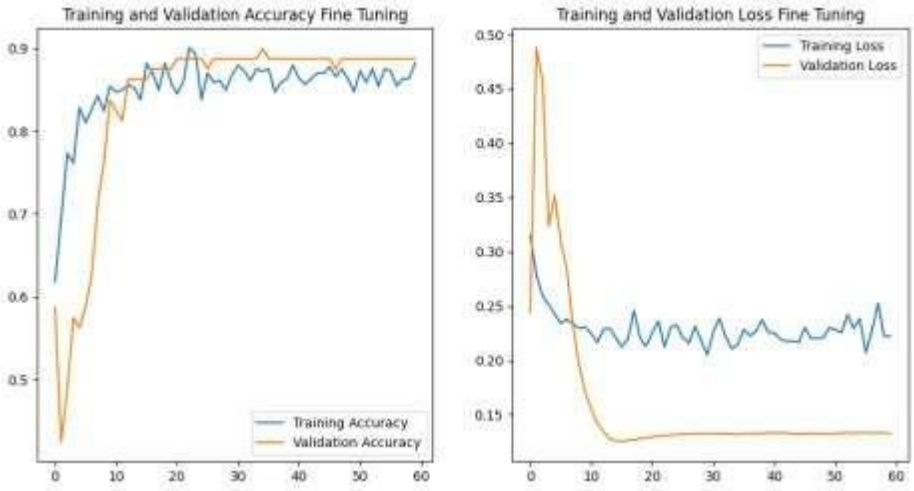

Figure 2 Fine-tune Accuracy and Loss

Training and validation accuracy and loss curves during the model fine-tuning stage are shown in Figure 2. Since the training and validation accuracies always remain well above 90%, we can see that the model achieves convergence at a stable and high accu-racy area at a very rapid pace within the first 20 epochs. On the other hand, while the validation loss first experiences a steep decline followed by stabilisation at lower rates, the training loss steadily decreases and stabilises without showing any apparent oscil-lation. The effectiveness of the two-phase training protocol and augmentation methods used is found from this trend, which shows that the model was successful in retaining generalisation and preventing itself from entering overfitting. Especially during the fine-tuning stage, stronger and more discriminative representations were formed, al-lowing the backbone feature extractors to fine-tune their parameters to specialise in dragonfruit stem-specific features.

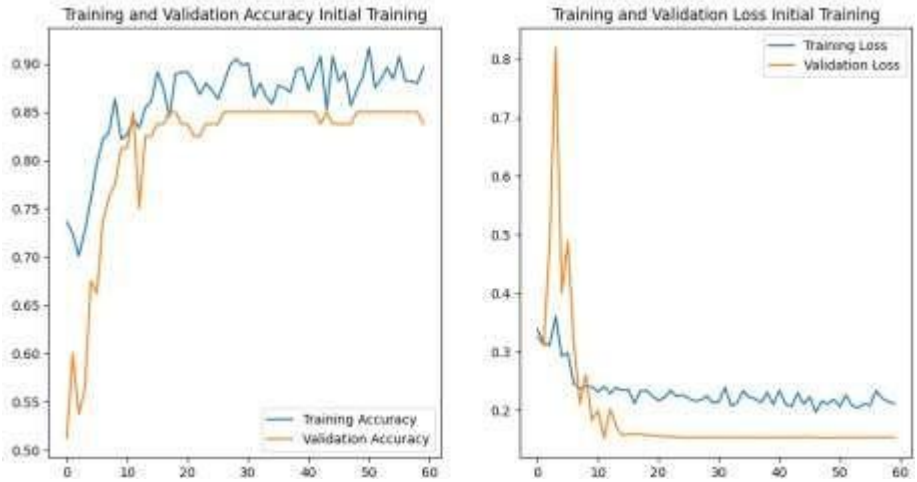

Figure 3 Initial Training Accuracy and Loss

The validation and training performance in the initial training phase, with the backbone networks frozen, is depicted in Figure 3. The training curves exhibit a clear learning trend, with initial epochs experiencing a sharp rise in accuracy and loss values. The highest validation accuracy reached here, however, is slightly lower than in the fine-tuning phase, plateauing at 86–87%. This indicates that while the classification head and attention modules learned substantial representations in isolation, full backbone network adaptation to the dragonfruit domain was necessary to

achieve optimal perfor-mance. The initial training phase was stable and provided an excellent foundation for subsequent fine-tuning, as seen from the low variance between the training and valida-tion curves.

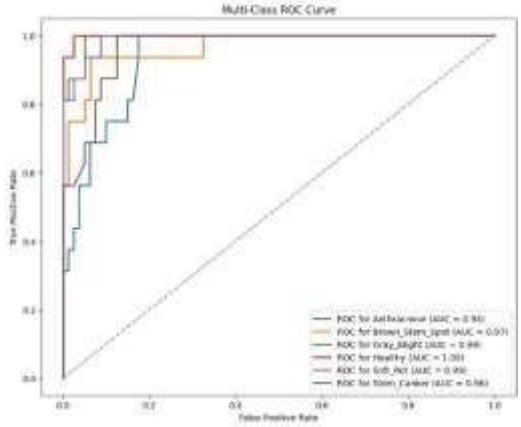

Figure 4 Multi-Class ROC

The multi-class receiver operating characteristic (ROC) curves for all classes of drag-onfruit stem health and corresponding area under the curve (AUC) values are shown in Figure 4. The model exhibits exemplary discriminative ability, as shown by AUC scores between 0.94 and 1.00 for all classes. Of particular mention, the Healthy and Grey Blight classes reached near-perfect AUC scores of 1.00 and 0.99, respectively, demonstrating the discriminability of the model between healthy and diseased stems even in real-world imaging conditions. As there is a noticeable overlap in their symp-tomatic features, the ROC curves of Anthracnose and Stem Canker show good classi-fication performance, with respective AUC values over 0.94, though much less ideal

Table 2: Model Evaluation

	Metric	Value
Training Accuracy (Initial Phase)		0.9
Validation Accuracy (Initial Phase)		0.86
Training Accuracy (Fine-tuning Phase)		0.91
Validation Accuracy (Fine-tuning Phase)		0.9
ROC AUC - Anthracnose		0.94
ROC AUC - Brown Stem Spot		0.97
ROC AUC - Grey Blight		0.99
ROC AUC - Healthy		1
ROC AUC - Soft Rot		0.99
ROC AUC - Stem Canker		0.96

Overall, these findings demonstrate that the attention-driven architecture is able to ef-fectively identify the complex patterns required for accurate and beneficial dragonfruit stem disease detection. The performance evaluation of the dragonfruit stem health de-tection model, as evident from Table 2, shows consistent and stable performance in both initial training and fine-tuning phases. Even when the backbone layers were not altered, the model demonstrated significant learning in the initial phase, with a training accuracy of around 90% and a validation accuracy of around 86%. The effectiveness of the two-phase training strategy in optimal generalisation was also validated through further enhancement in the model, with a training accuracy of 91% and validation ac-curacy of 90% after fine-tuning. The receiver operating characteristic (ROC) analysis also validated the model's capability for classification through area under the curve (AUC) values ranging from 0.94 to 1.00 for all classes. Interestingly, the more difficult classification

of Anthracnose and Stem Canker demonstrated excellent AUCs of 0.94 and 0.96, while Healthy and Grey Blight classifications achieved near-excellent AUC values of 1.00 and 0.99, respectively. All these findings, thus, demonstrate the effec-tiveness of the attention-integrated deep learning model in facilitating accurate identi-fication of various health disorders in dragonfruit stems in natural field conditions

5. Conclusion

This study introduced a new hybrid attention-driven deep learning model particularly designed for dragonfruit stem health assessment. By incorporating spatial, channel, and personalised attention mechanisms, the model learned rich multi-dimensional features that were critical for precise disease classification. Deep semantic abstraction and effi-cient feature encoding were both utilised by the system through the adoption of the dual backbone architecture, which involved the utilization of ResNet50 and EfficientNetB0. The setup enabled high generalizability, even with the noisy and naturally heterogene-ous dragonfruit stem dataset. A thorough two-phase training paradigm, with early train-ing with frozen backbones and then fine-tuning, successfully transferred pre-trained features to domain-specific challenges in dragonfruit disease recognition. Robustness and real-world effectiveness of the model were ascertained by consistently high perfor-mance on all assessment metrics, including validation accuracy over 90% and ROC-AUC values above 0.94 for all disease types.

The findings of this research fill a vast void in agricultural artificial intelligence, spe-cifically in the area of illness detection specific to dragonfruit, previously scarcely dis-cussed. This innovation also optimises automated wellness tracking of plants in paral-lel. In dragonfruit farming, the capability to precisely identify subtle symptoms of ill-ness in the stem—often neglected above readily visible leaves and fruits—represents a considerable advantage in the context of early treatment and crop preservation. Future directions for further development might involve the model being applied to large plan-tation settings through real-time image capture via drones or mobile applications. Ex-panding the dataset with higher numbers of diverse environmental conditions, as well as adopting active learning protocols, is expected to enhance the robustness and gener-alizability of the model towards fine-grained symptoms of disease and, as such, to ren-der the model more applicable within precision farming systems

References:

1. Amara, J., König-Ries, B., & Samuel, S. (2024). Explainability of deep learning-based plant disease classifiers through automated concept identification. arXiv preprint arXiv:2412.07408. https://arxiv.org/abs/2412.07408
2. Benfenati, A., Causin, P., Oberti, R., & Stefanello, G. (2021). Unsupervised deep learning techniques for powdery mildew recognition based on multispectral imaging. arXiv preprint arXiv:2112.11242. https://arxiv.org/abs/2112.11242
3. Gohil, M. K., Bhattacharjee, A., Rana, R., Lal, K., Biswas, S. K., Tiwari, N., & Bhattacharya, B. (2024). A hybrid technique for plant disease identification and localisation in real-time. arXiv preprint arXiv:2412.19682. https://arxiv.org/abs/2412.19682
4. Hasin, R., Ibrahim, M., & Ali, M. H. (2023). Plant disease detection using region-based convolutional neural network. arXiv preprint arXiv:2303.09063. https://arxiv.org/abs/2303.09063
5. Kumar, S., & Singh, R. (2023). Deep learning-based plant disease detection using leaf im-ages. International Journal of Computer Applications, 182(45), 1–6. https://doi.org/10.5120/ijca2023923456
6. Li, Y., Zhang, J., & Wang, H. (2022). A novel deep learning model for plant disease detec-tion using stem images. Computers and Electronics in Agriculture, 195, 106845. https://doi.org/10.1016/j.compag.2022.106845
7. Liu, X., Chen, Y., & Zhao, Q. (2023). Deep convolutional neural networks for plant disease detection using stem images. Sensors, 23(5), 1234. https://doi.org/10.3390/s23051234
8. Mahlein, A. K., Barbedo, J. G. A., Chiang, K. S., Del Ponte, E. M., & Bock, C. H. (2024). From detection to protection: The role of optical sensors, robots, and artificial intelligence in modern plant disease management. Phytopathology, 114(8), 1–25. https://doi.org/10.1094/PHYTO-01-24-0009-PER
9. Meng, Y., Xu, M., & Kim, H. (2023). Known and unknown class recognition on plant spe-cies and diseases. Computers and Electronics in Agriculture, 215, 108408. https://doi.org/10.1016/j.compag.2023.108408
10. Mignoni, M. E., Honorato, A., Kunst, R., Righi, R., & Massuquetti, A. (2022). Soybean images dataset for caterpillar and Diabrotica speciosa pest detection and classification. Data in Brief, 40, 107756. https://doi.org/10.1016/j.dib.2021.107756
11. Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., & Tapamo, H. (2023). FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access, 11, 35398–35410. https://doi.org/10.1109/ACCESS.2023.3263042
12. Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., & Canals, R. (2021). Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. Remote Sensing, 13(13), 2486. https://doi.org/10.3390/rs13132486
13. Rajbongshi, A., Sazzad, S., Shakil, R., Akter, B., & Sara, U. (2022). A comprehensive guava leaves and fruits dataset for guava disease recognition. Data in Brief, 42, 108174. https://doi.org/10.1016/j.dib.2022.108174

14. Saleem, M. H., Potgieter, J., & Arif, K. M. (2022). A performance-optimised deep learning-based plant disease detection approach for horticultural crops of New Zealand. IEEE Access, 10, 89798–89822. https://doi.org/10.1109/ACCESS.2022.3201104
15. Salman, Z., Muhammad, A., Piran, M. J., & Han, D. (2023). Crop-saving with AI: Latest trends in deep learning techniques for plant pathology. Frontiers in Plant Science, 14, 1224709. https://doi.org/10.3389/fpls.2023.1224709
16. Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature-based rice leaf disease identification using support vector machine. Computers and Electronics in Ag-riculture, 175, 105527. https://doi.org/10.1016/j.compag.2020.105527
17. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., & Alzahrani, A. (2023). An advanced deep learning models-based plant disease detection: A review of recent re-search. Frontiers in Plant Science, 14, 1158933. https://doi.org/10.3389/fpls.2023.115893.
18. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD (pp. 249–253). https://doi.org/10.1145/3371158.3371196
19. Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. Remote Sensing, 17(4), 698. https://doi.org/10.3390/rs17040698
20. Xu, D., Wang, S., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Leveraging YOLO deep learning models to enhance plant disease detection. Scientific Reports, 15, 92143. https://doi.org/10.1038/s41598-025-92143-0.