Received: 5 May 2025 | Revised: 15 August 2025 | Accepted: 17 October 2025 | Published online: 17 December 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1-10
DOI: 10.47852/bonviewAIA52026081

BON VIEW PUBLISHING

Adaptive Swin Transformer V2-Tiny Based Model for Classification of Bacteria, Fungus, Virus, and Healthy Fruit and Leaf Images

Poornima Basatti Hanuma Gowda[1], Basavanna Mahadevappa[1,*], Shivakumara Palaiahnakote[2,3], Muhammad Hammad Saleem[2,3], and Niranjan Mallappa Hanumanthu[4]

[1] *Department of Studies in Computer Science, Davangere University, India*

[2] *School of Science, Engineering, and Environment, University of Salford, United Kingdom*

[3] *Data Science and Artificial Intelligence Hub, University of Salford, United Kingdom*

[4] *Department of Studies in Biotechnology, Davangere University, India*

Abstract: The classification of fruits and leaves affected by bacteria, viruses, and fungi has made significant progress in the fields of artificial intelligence and image processing. However, most methods focus on particular categories of fruit and leaf diseases, but not on both fruit and leaf diseases caused by bacteria, viruses, and fungi. This study aimed to develop a model for the classification of the initial, intermediate, and final stages of bacterial, viral, and fungal diseases, irrespective of fruit and leaf types. To achieve this goal, inspired by the accomplishments of the Swin Transformer, the Swin Transformer V2-Tiny was explored for the classification of 10 classes, which included healthy and three stages of bacteria, virus, and fungus images of fruits and leaves. The stages of Swin Transformer V2-Tiny divide the image into patches, namely, linear projection, Window Multi-Head Self-Attention (W-MSA), and Shifted Window Multi-Head Self-Attention (SW-MSA) for local and global features, which were adapted to perform the plant disease classification. Experiments on authors' curated and standard datasets and a comparative study with recent methods demonstrate effective classification and superiority over existing methods. To the best of our knowledge, this is the first study on the classification of fruit and leaf pathogens caused by bacteria, viruses, and fungi based on their development stages. The proposed model achieved an average classification rate of 91.04% on fruit datasets and 94.07% on leaf datasets, outperforming recent benchmark methods. It also demonstrated strong generalization on unseen public datasets with over 93% accuracy.

Keywords: fruit classification, leaf classification, fruit/leaf disease classification, Swin Transformer

1. Introduction

Plant diseases caused by bacterial, fungal, and viral infections pose a significant threat to agricultural productivity. Therefore, the early detection of these diseases at an early stage is crucial for implementing timely interventions and reducing crop losses. In addition, it reduces manpower and expenses, leading to cost effectiveness. Plant, fruit, and leaf disease identification is not a new challenge, and we can find several existing methods can be found in the literature [1–3]. Existing methods use Convolutional Neural Networks (CNNs) and other deep learning models.

However, as existing methods focus on a particular dataset and disease, they are not robust enough to handle the stages of fruit and leaf diseases caused by bacteria, viruses, and fungi. For example, sample fruit and leaf images of the initial, intermediate, and final stages caused by bacteria, viruses, and fungi are shown in Figures 1(a)–(i) and 2(a)–(i). It is observed from Figures 1(a)–(i) and 2(a)–(i) that the difference between samples of different diseases is minimal. This is because the images share common observations. The same is true for the leaf images. Although observations such as yellow patches for viruses, dark patches for bacteria, and white patches for fungi can differentiate diseases at different levels, unpredictable shapes, structures, and varieties make the classification problem more complex and challenging. Thus, the classification of disease pathogens at different levels, irrespective of fruit and leaf types, is an open challenge.

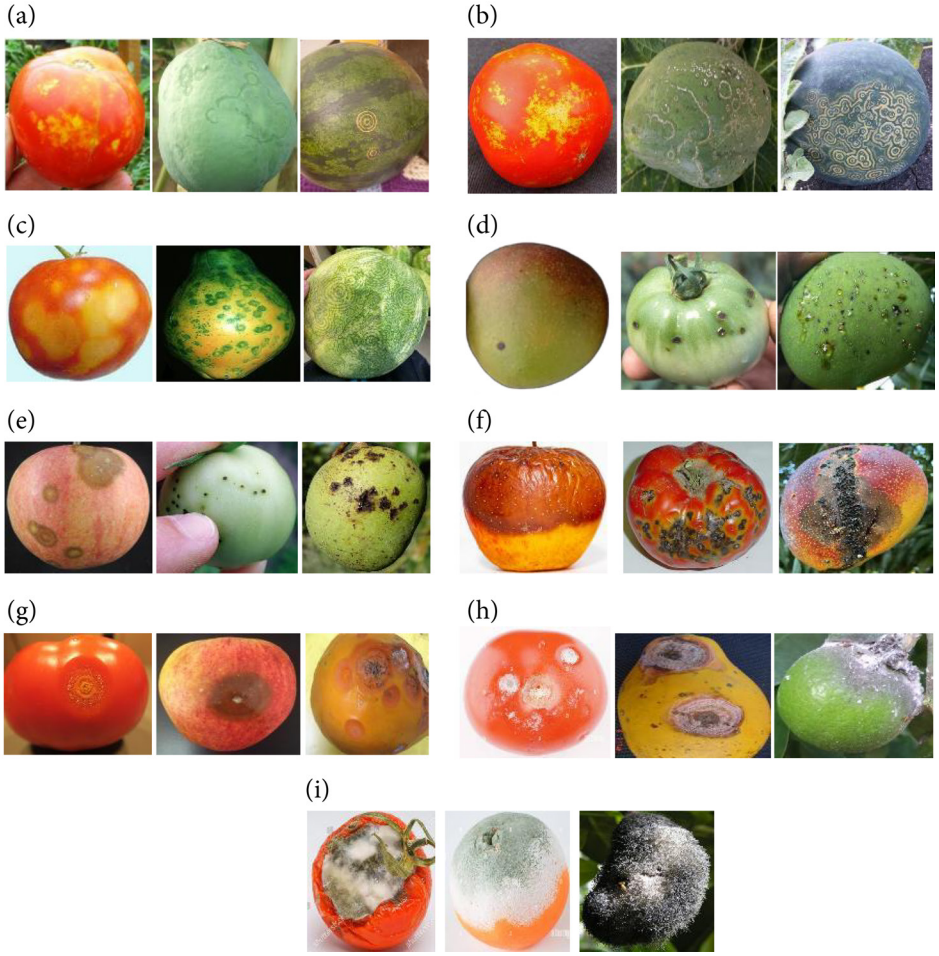
To address this challenge, this study explored the Swin Transformer V2-Tiny model for the classification of fruit and leaf images of bacteria, viruses, and fungi at initial, intermediate, and final levels. Figures 1 and 2 show that the observations, namely, yellow, dark, and white patches for viral, bacterial, and fungal infections, respectively, are vital cues for successful classification. To extract such observations, inspired by the elegant vision transformer, which is a special model for visual feature extraction, the proposed method adapts the Swin Transformer V2-Tiny [4–6]. Compared to the baseline, the Swin Transformer, the Swin Transformer V2-Tiny is robust, easy to adapt to different situations, and extracts features efficiently. This motivated us to explore the Swin Transformer V2-Tiny in contrast to the baseline Swin Transformer for the classification of healthy and diseased pathogens of plants.

The key contributions of this study are as follows:

1) Exploring the Swin Transformer V2-Tiny model for the classification of bacteria, virus, and fungus-infected fruit and leaf images at the initial, intermediate, and final levels.

***Corresponding author:** Basavanna Mahadevappa, Department of Studies in Computer Science, Davangere University, India. Email: basavanna.m@davangereuniversity.ac.in

Figure 1
Sample images of virus, bacteria, and fungus at initial, intermediate, and final stages of fruits



2) Adapting the Swin Transformer V2-Tiny for the classification of infected images, irrespective of fruit and leaf type.

The remainder of this paper is organized as follows: Section 2 discusses related work on fruit and leaf disease classification and identification. Section 3 presents details of the proposed methodology, including the model architecture. Section 4 presents the experimental results and evaluation metrics. Section 5 discusses the findings and limitations of the study.

2. Related Work

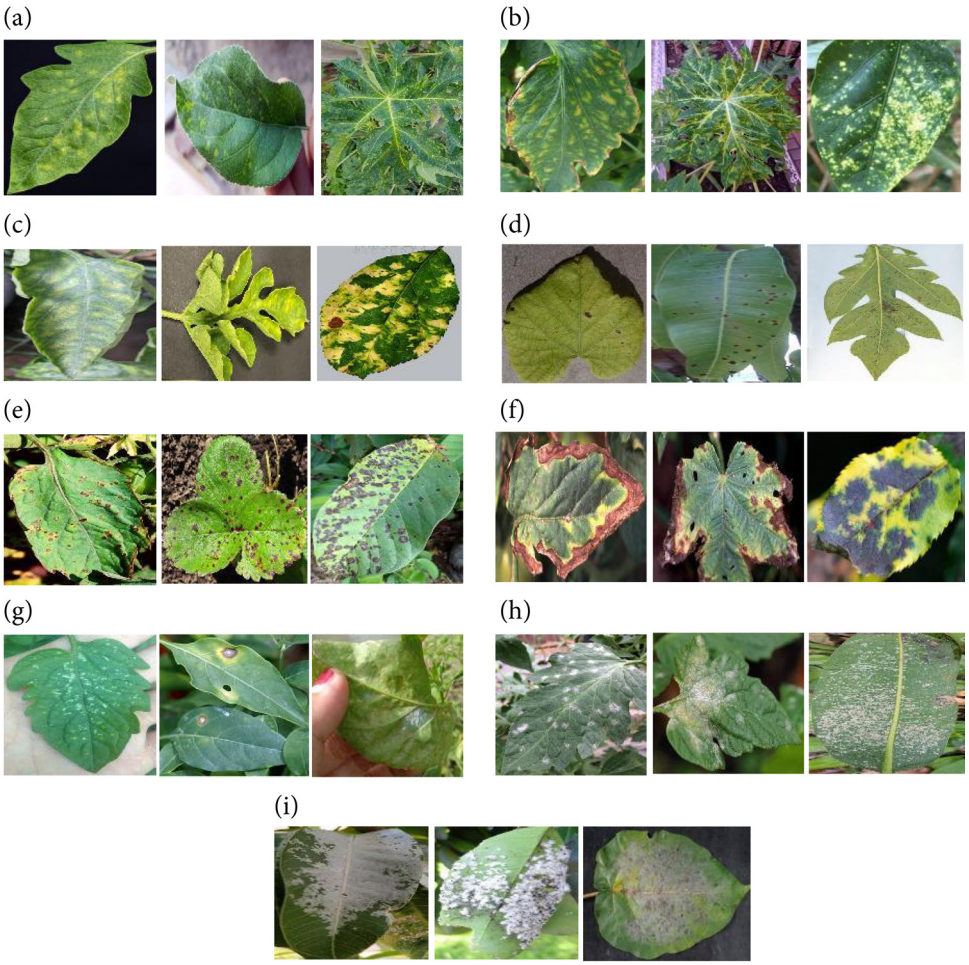
If we consider the classification of bacteria, fungi-, and virus-infected images as a general image classification problem, several methods are available in the literature. For instance, the methods [4, 7] were proposed for object detection and classification in images. Because objects in images have unique shapes, these existing methods extract features that represent the shapes of particular objects. However, in our classification, the patches did not have any shape, and the images had unpredictable shapes. Therefore, general image classification methods may not work well for the classification of virus-, bacteria-, and fungus-infected images based on their development stages.

2.1. General image classification

Li et al. [4] introduced PMST (Parallel and Miniature Swin Transformer), a refined variant of the Swin Transformer tailored for logo detection. By incorporating a bypass-parallelizable shift module and a miniature window tandem shift strategy, the model enhances feature fusion and information transfer between windows, addressing challenges such as varying logo scales, diversity, and distortions in the data. Similarly, He et al. [7] proposed Dual-branch Swin Transformer with Asymmetric Attention Fusion (DST-A2F) for radar-based gait recognition. This model extracts distinctive features from spectrogram and cadence velocity diagram (CVD) representations, prioritizing more informative spectrogram features using asymmetric attention fusion.

Furthermore, Hu et al. [8] developed MAFDet, a multi-attention fusion network for small-object detection in drone imagery. This method combines a Swin Transformer backbone, a multi-attention focusing sub-network, and an anchor-free detection head, significantly improving detection accuracy on datasets such as VisDrone and UAVDT. Meanwhile, Hu et al. [5] presented VGG-TSwinformer, a deep learning framework integrating VGG-16 and Swin Transformer for early Alzheimer's disease (AD) diagnosis using longitudinal MRI data. Leveraging spatial and temporal attention mechanisms for classification. Zeng et al. [6] introduced DTMINet, a Dual Swin-Transformer-based Mutual Interactive Network for RGB-D salient object detection, incorporating novel fusion mechanisms and outperforming state-of-the-art techniques. Similarly, Zhang and Tu [9] proposed SwinFR, a super-resolution model for remote sensing images that combines SwinIR with Fast Fourier Convolution (FFC) to enhance low-frequency information retention. Finally, Zhou et al. [10] presented an improved YOLOv7 model incorporating Modulated Deformable Convolution and Swin Transformer to enhance object detection in fisheye images, achieving superior performance on VOC-360 and ERP-360 datasets. Pal et al. [11] provided a comprehensive review of text detection and

Figure 2
Sample images of virus, bacteria, and fungus at initial, intermediate, and final stages of leaves



recognition methods in natural scene images, discussing various deep learning approaches, including regression-based, segmentation-based, and transformer-based models. This study highlights the strengths and limitations of these techniques and identifies open challenges in handling complex scenarios, such as low-light conditions and arbitrary motion.

Deep learning and ML-based classification methods have been widely used in the medical and agricultural domains. Kang et al. [12] successfully implemented machine learning models for COPD classification, achieving high accuracy using decision-tree-based classifiers. This methodology is relevant for plant disease classification, where similar ML techniques can be used to distinguish between bacterial, fungal, and viral infections.

Despite the advancements in Swin Transformer applications across various fields, existing image classification techniques are not designed to classify multiple domains, such as fruit and leaf infections at different levels, caused by bacteria, fungi, and viruses. Therefore, they may not be effective for specific tasks.

2.2. Fruit and leaf disease classification

Gupta and Tripathi [13] conducted a comprehensive survey analyzing fruit and vegetable disease classification methods using machine learning, deep learning, and IoT-based technologies. Similarly, Laim et al. [14] developed an automated fruit disease classification model using the Scale-Invariant Feature Transform (SIFT), which extracts key image features to enhance classification accuracy. Aboelenin et al. [15] proposed a hybrid deep learning framework integrating CNN and Vision Transformers (ViT) to detect plant leaf diseases, achieving state-of-the-art accuracy on apple and corn leaf disease datasets.

Other notable works include Xu et al. [16] who introduced PDNet, a parameter-efficient Vision Transformer model designed for plant leaf disease identification. Their method utilized Overlapping Patch Embedding (OPE) and Angular Softmax Loss (A-Softmax) for refined disease classification, demonstrating superior performance across multiple datasets. Additionally, Megalingam et al. [17] developed a Vision Transformer-based approach for cowpea leaf disease classification, achieving 96% accuracy. Further contributions by Das et al. [18] provided an extensive review of deep learning techniques for tomato leaf disease classification, whereas Wu et al. [19] introduced an attention-based CNN (LBPAttNet) to improve tea leaf disease identification.

2.3. Plant disease classification

Barman et al. [20] proposed ViT-SmartAgri, a Vision Transformer-based model for plant disease detection, achieving 90.99% accuracy on

smartphone-captured tomato leaf images. Hemalatha and Jaychandran [21] introduced PDLC-ViT, a multi-task Vision Transformer model for disease localization and classification, achieving state-of-the-art performance on the Plant Village and PlantDoc datasets. Tunio et al. [22] developed a Transformer-fused CNN model using Wasserstein domain adaptation to improve generalization across plant disease datasets. Additionally, Singh et al. [23] leveraged synthetic data augmentation using LeafyGAN to enhance plant disease classification performance. Liu and Zhang [24] introduced an Efficient Swin Transformer that integrates selective token generation and feature fusion to improve accuracy and reduce computational complexity. Subramanian et al. [25] compared conventional and AI-driven intervention techniques for Alzheimer's disease. They discuss traditional methods such as cognitive stimulation and reminiscence therapy, alongside AI-based approaches such as deep learning, vision transformers, and NLP for early diagnosis and personalized care. The study highlights the potential of integrating both approaches to enhance the management of patients with Alzheimer's disease.

Balasundaram et al. [26] applied the Segment Anything Model (SAM) in conjunction with deep convolutional neural networks for tea-leaf disease detection, demonstrating improved segmentation-guided classification performance under field conditions.

Hasan et al. [27] introduced a comprehensive smartphone image dataset of radish plant leaves from Bangladesh to support automated disease classification. This dataset enables effective benchmarking of deep learning models and contributes to enhancing model generalization in real-world agricultural conditions.

Ni et al. [28] demonstrated the use of FTIR (Fourier-transform infrared) spectroscopy in combination with machine-learning methods to diagnose corn leaf diseases, providing a spectral-based alternative to conventional image-based classification approaches.

Petchiammal and Murugan [29] explored automated identification of paddy leaf diseases using visual leaf images and compared nine pre-trained deep-learning architectures (including VGG16, VGG19, DenseNet variants, MobileNetV2, InceptionV3 and ResNet152V2) to evaluate classification performance across multiple disease classes.

Akhter and Saxena [30] documented for the first time the co-infection of papaya in Lucknow (India) by Catharanthus yellow mosaic virus (CaYMV) together with a novel betasatellite they named Tomato leaf curl Lucknow betasatellite (ToLCLB), expanding understanding of the complex begomovirus–satellite combinations underlying Papaya leaf curl disease (PaLCD) in Indian papaya.

Although these methods focus on plant disease identification, they may not generalize well when classifying multiple types of diseases in different plant species.

2.4. Virus, bacteria, and fungus disease classification

Saraswat et al. [1] developed an advanced method for detecting fungal and bacterial diseases in plants using a modified deep neural network (MDNN) combined with the Dynamic SURF (DSURF) technique. Gaikwad et al. [2] introduced a deep CNN model for fungi-affected fruit leaf disease classification using AlexNet and SqueezeNet. Furthermore, Gaikwad et al. [3] focused on CNN-based identification of fungi-infected guava leaves using SqueezeNet. Other contributions include those of Priyanka et al. (2025) who analyzed papaya leaf curl disease (PaLCD) caused by begomoviruses and identified novel viral strains in India. Pakruddin and Hemavathy [31] introduced a pomegranate disease dataset to aid the deep learning-based classification of bacterial blight, anthracnose, and Alternaria fruit spot. Additionally, Siripatrawan and Makino [32] utilized hyperspectral imaging (HSI) with machine learning techniques to detect anthracnose in mangoes, demonstrating the effectiveness of presymptomatic detection.

Table 1
Comparative overview of recent models and their performance in plant disease classification

Study	Model/Method	Dataset(s)	Key techniques	Advantages	Limitations/Research gaps
Aboelenin et al. [15]	CNN + ViT Hybrid	Apple & Corn Leaf Datasets	CNN for feature extraction, ViT for classification	High accuracy, better generalization than CNN alone	Dataset-specific, lacks stage-wise disease classification
Xu et al. [16]	PDNet (Transformer)	Multiple Plant Leaf Datasets	Overlapping Patch Embedding (OPE), A-Softmax loss	Efficient transformer design, fewer parameters	Does not address multi-stage disease progression
Tunio et al. [22]	Transformer-fused CNN + Domain Adaptation	PlantVillage, PlantDoc	Wasserstein domain adaptation, hybrid fusion	Good cross-domain performance	No specialization for bacteria, virus, fungus stages
Singh et al. [23]	ViT + GAN (LeafyGAN)	Tomato Leaf Dataset	Synthetic image generation, transformer-based classifier	Works well with limited real data	Not generalized across multiple crops/diseases
Hu et al. [9]	MAFDet (Transformer-based detector)	Drone imagery datasets	Multi-attention fusion, anchor-free detection	Effective for small object detection	Not focused on disease classification
Liu and Zhang [24]	Efficient Swin Transformer	Custom Plant Disease Dataset	Selective token generation, lightweight Swin blocks	Reduced complexity, fast inference	Limited to detection, not detailed classification
Das et al. [18]	Deep CNNs (Survey)	Tomato Leaves	Review of multiple deep learning methods	Highlights state-of-the-art models	Does not provide a unified solution
Barman et al. [20]	ViT-SmartAgri	Smartphone images (Tomato)	Vision Transformer, mobile deployment	Suitable for real-time field applications	Single crop, lacks stage-wise analysis
This Work	Swin Transformer V2-Tiny (Modified)	Custom 10-Class + NZDL-v1/v2	Local-global attention (W-MSA + SW-MSA), stage-wise classification, modified classifier head	Handles multiple diseases, all stages, high accuracy, generalizable	None—addresses disease type + stage classification in fruit and leaf images

While these studies address virus, bacteria, and fungus disease identification in specific crops, none have focused on a unified classification model for all three types. The lack of domain-independent methods indicates that existing techniques are designed for specific scenarios, limiting their applicability to diverse datasets.

Recent research has demonstrated the effectiveness of deep learning in plant disease classification, with increasing attention on Vision Transformers (ViT) and hybrid CNN-ViT architectures due to their superior feature extraction capabilities. For instance, Aboelenin et al. [15] proposed a hybrid CNN-ViT model achieving high accuracy in detecting apple and corn leaf diseases. Xu et al. [16] introduced a parameter-efficient ViT model using Overlapping Patch Embedding and A-Softmax loss, enabling accurate disease identification with fewer parameters. Similarly, Tunio et al. [22] developed a transformer-fused CNN with Wasserstein domain adaptation, which improved generalization across plant disease datasets. Singh et al. [23] applied GAN-based synthetic augmentation to train vision transformers more effectively on limited real data. While these methods highlight the power of transformers, they are often tailored to specific crops, do not generalize across disease stages, or overlook the early-stage variations critical for effective disease control.

Table 1 presents a comparative summary of the recent models and their effectiveness in plant disease classification.

In summary, from the literature review reported in Table 1, it is evident that no existing models simultaneously classify images of bacteria-, fungi-, and virus-infected fruit and leaves at different levels. Consequently, there is a pressing need for an effective classification model to address this issue. Hence, this study aims to develop such a model to bridge this gap and provide a more comprehensive solution for disease classification. Unlike existing models, the proposed Tiny Swin Transformer V2 architecture is adapted to handle subtle inter-class differences and high visual similarity across stages, making it suitable for real-world, diverse datasets.

3. Proposed Methodology

As noted in the sample images in Figures 1 and 2, the yellow, dark, and white patches are the key cues for differentiating fruit and leaf images infected by viruses, bacteria, and fungi at different levels. These observations were made by experts in the Biotechnology Department. It was also observed that the size and thickness of the patches changed as the levels changed (initial, intermediate, and final). These observations motivated us to introduce a vision transformer to extract the features that represent unique observations. Because the difference in the images at different levels is marginal, the baseline vision transformer may not be effective. Therefore, a robust Tiny Swin Transformer V2 model was used for successful classification. Compared with the classical Swin Transformer model, the Tiny Swin Transformer V2 is efficient, accurate and adaptable to different situations. Hence, the proposed Method was explored for the classification of stages of viruses, bacteria, and fungi of fruits and leaves. The block diagram of the proposed method is shown in Figure 3.

Figure 3
The block diagram of the proposed method

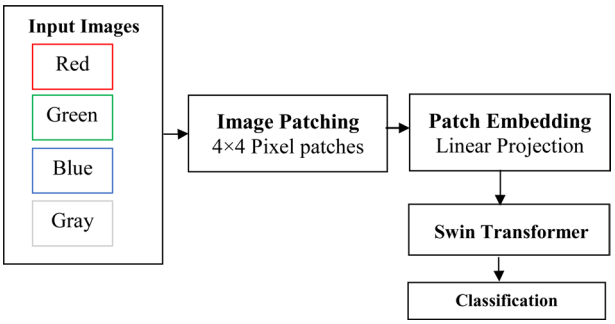


Figure 3 shows the input images of different colors. The input images were divided into patches and projected linearly to extract features. The Swin Transformer extracts deep hierarchical features from images. The extracted features were fed into the classification step to classify the initial, intermediate, and final stages of virus, bacteria, and fungus, along with a healthy class.

3.1. Classification

The Tiny Swin Transformer V2 model, a hierarchical vision transformer, enhances feature extraction through its shift-window mechanism, allowing it to efficiently capture both local and global dependencies in images. Local Dependencies refer to short-range relationships between neighboring patches or pixels in an image. These capture fine-grained details, such as texture, edges, or small variations in color, which are crucial for distinguishing subtle differences, especially in early-stage disease symptoms. Global Dependencies refer to long-range relationships between distant regions of the image. These help in understanding the overall structure or patterns that span across larger parts of the image, such as widespread lesions or disease spread patterns. The proposed model applies transfer learning using Tiny Swin Transformer V2, which was trained on augmented fruit and leaf image data.

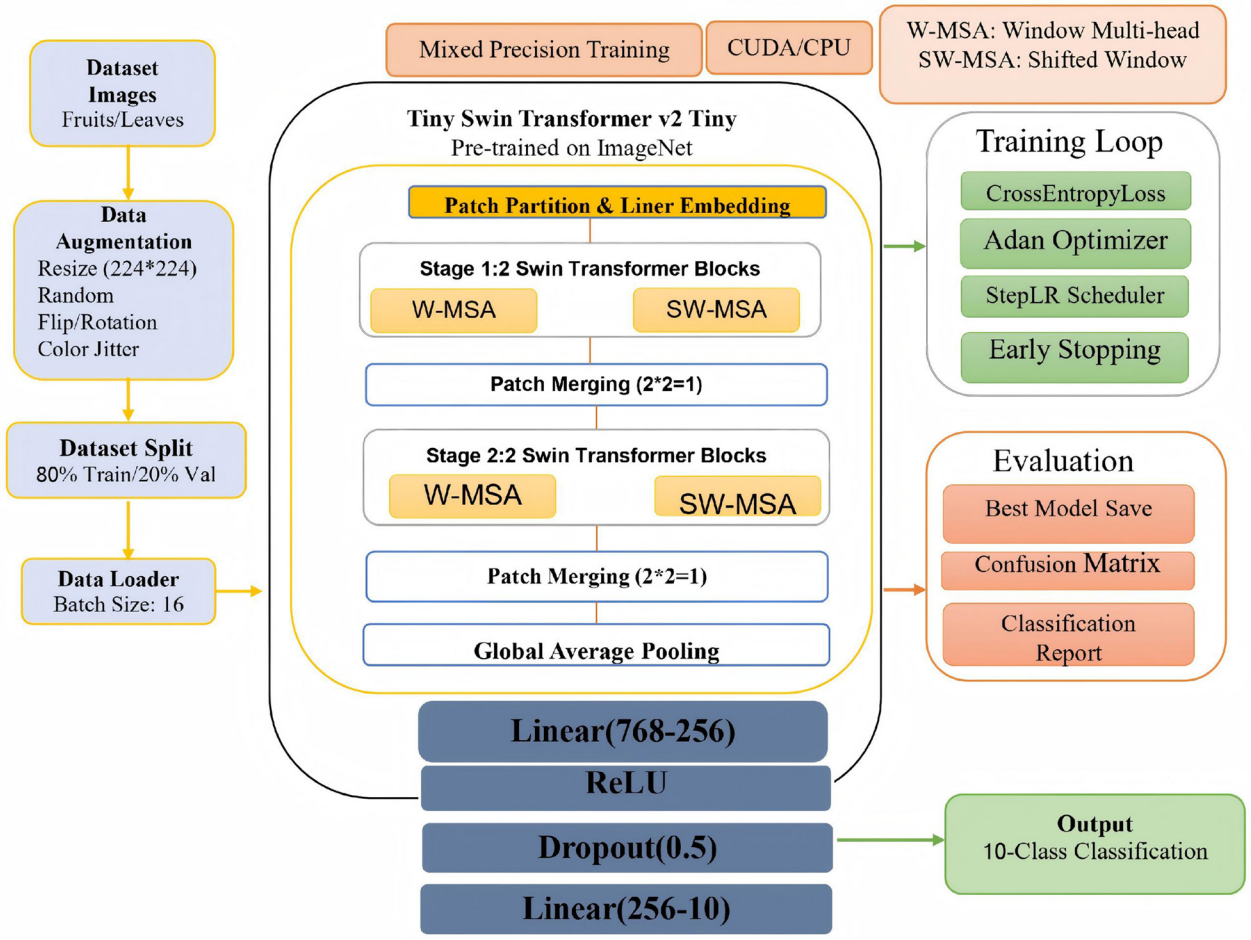
This 10-class classification framework utilizes the Tiny Swin Transformer V2 model, as shown in Figure 4, a hierarchical vision transformer pre-trained on ImageNet, to classify fruit and leaf diseases caused by viruses, bacteria, and fungi at different infection stages (initial, intermediate, and final). The pipeline begins with RGB dataset images, which undergo data augmentation techniques such as resizing (224 × 224) to match the input size of the model, random flipping and rotation to introduce geometric variability, and color jittering to enhance robustness against lighting variations. The images were then normalized using ImageNet statistics and split into 80% training and 20% validation sets, ensuring balanced class representation. A PyTorch DataLoader was configured with a batch size of 16 to efficiently load the images during training. The Swin Transformer processes input images by first partitioning them into non-overlapping 4 × 4 patches, followed by a linear embedding that transforms each patch into a 96-dimensional feature vector.

The feature extraction pipeline consists of two hierarchical stages, each containing two Swin Transformer blocks that utilize Window Multi-Head Self-Attention (W-MSA) and Shifted Window Multi-Head Self-Attention (SW-MSA) to capture both local and global dependencies. Between stages, the patch merging layers (2 × 2→1) progressively reduced the spatial dimensions while increasing the feature depth, allowing the model to learn hierarchical representations. After the final stage, Global Average Pooling (GAP) was applied to compress the spatial information into a single feature vector. The classification head consisted of a fully connected layer (Linear 768→256), followed by a ReLU activation, Dropout (0.5) to prevent overfitting, and another fully connected layer (Linear 256→10) that mapped the features to 10 class logits. The model was trained using mixed precision (AMP) on CUDA or CPU, optimizing with CrossEntropyLoss, the Adam optimizer, a StepLR scheduler (reducing the learning rate every 10 epochs by a factor of 0.1), and early stopping (patience = 5 epochs) to prevent overfitting.

More details of the proposed method are presented below.

Patch Partitioning and Linear Embedding: The input RGB images are resized and partitioned into non-overlapping patches as mentioned earlier, resulting in a grid of 56 × 56 patches. Each patch is flattened and passed through a linear projection to create a

Figure 4
Swin Transformer Tiny-V2 architecture for classification of fruit and leaf



96-dimensional embedding vector. These embeddings form the input to the Swin Transformer blocks. Although Tiny Swin Transformer V2 begins by dividing the image into non-overlapping 4 × 4 patches, it does not reassemble these patches into the original image in a spatial sense. Instead, it preserves and reconstructs the global context through a combination of two mechanisms:

Shifted Window Attention (SW-MSA): By shifting the attention window across layers, the model ensures that each patch interacts with its neighboring patches from adjacent windows. This overlapping attention helps bridge patch boundaries and recovers spatial relationships. Hierarchical Feature Aggregation: As the model proceeds through its stages, the patch merging operations combine adjacent patches and increase the receptive field. This forms a deep, multi-scale representation of the entire image, allowing the model to reason about both local and global structures. These mechanisms allow the model to encode the entire image context without explicitly reassembling the patches, thus ensuring that no critical information is lost during the process.

Hierarchical Feature Extraction with Local and Global Dependencies: Swin Transformer V2-Tiny employs a hierarchical architecture with four stages. W-MSA enables the model to capture local dependencies by applying self-attention within non-overlapping windows of fixed size as discussed above. SW-MSA shifts the windows between layers, facilitating cross-window interactions and thereby capturing global dependencies without excessive computation. This mechanism allows the network to model subtle variations in patch shapes and textures across infection stages. Each stage is followed by a Patch Merging layer, where adjacent 2 × 2 patches are concatenated and passed through a linear layer. This reduces spatial resolution and increases channel depth, allowing the model to build deep hierarchical features from low-level edges to high-level semantic cues.

Modified Classifier Head: The default classifier head of the pre-trained Swin V2-Tiny model is replaced with a task-specific fully connected head: A linear layer reduces the feature dimension from 768 to 256 followed by a ReLU activation and Dropout (p = 0.5). Another linear layer maps 256-dimensional features to the final 10-class logits.

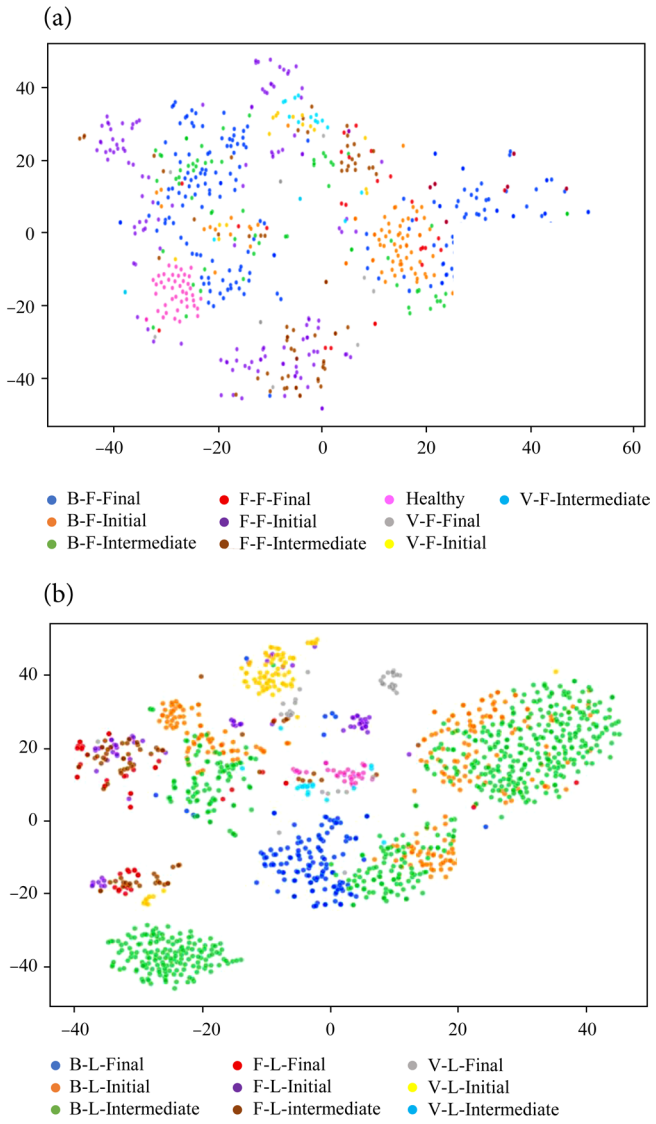
This modification enhances the model's capability to learn disease-specific class boundaries in the high-dimensional feature space.

Training Enhancements and Implementation Details: To ensure robustness, generalization, and reproducibility, the following training strategies were employed: Transfer Learning: We initialized the Tiny Swin Transformer V2 model with ImageNet-1K pretrained weights and fine-tuned it on our 10-class dataset.

Loss Function: CrossEntropyLoss with label smoothing (smoothing factor = 0.1) was used to prevent overconfidence and improve generalization. The model was trained with the Adam optimizer (learning rate = 3e−4, weight decay = 1e−5). A Step learning rate scheduler reduced the learning rate by a factor of 0.5 every 5 epochs, implemented with a patience of 10 epochs to prevent overfitting and reduce training time, and enabled via PyTorch's AMP (torch.cuda.amp) to accelerate training and reduce memory usage on CUDA-compatible GPUs with batch Size: 16, Epochs: 30, and Input Size: 224 × 224 pixels.

Data Augmentation: To increase model robustness and reduce overfitting, we applied comprehensive augmentation techniques,

Figure 5
(a) distribution of classes of fruits and (b) leaf disease classification



Note: B-F — Bacteria on fruit, F-F — Fungus on fruits, V-F — Virus on fruits, B-L — Bacteria on leaf, F-L — Fungus on leaf, V-L — Virus on leaf.

namely, Resize (224 × 224), Random Horizontal Flip, Random Rotation (±30°), Color Jitter (Brightness, Contrast, Saturation, Hue), Random Affine Transformations (translation ±10%), Normalization with ImageNet Statistics with Mean = [0.485, 0.456, 0.406] and Std = [0.229, 0.224, 0.225].

To visualize the effectiveness of the proposed model on classification, the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm was used to map high-dimensional data to a lower-dimensional space, as shown in Figure 5(a)–(b) for fruits and leaves. The two t-SNE plots represent the feature embeddings of fruits and leaves in Figure 5(a) and (b), where it is noted that the Swin Transformer V2-Tiny t-SNE plot distinguishes almost all the classes with certain overlapping for both fruit and leaf images. Owing to some overlap, the performance of the proposed model degraded. This shows that this classification of healthy and diseased fruits and leaves is a complex problem. This indicates that there is scope for further improvement.

4. Experimental Results

We collected a dataset from the Biotechnology Department of Davangere University, Karnataka, India, to evaluate the proposed and existing methods. The main problem is that the images must be manually labelled for experimentation. However, this manual process requires more time; hence, it is difficult to handle a large variety of images affected by different diseases using this method. The main goal of this study is to integrate the proposed system with their devices to assist their investigation. To make the collection as comprehensive and representative as possible, images were collected from multiple sources, areas, fields, and open spaces under various weather conditions. Sample images illustrating the wide diversity of inputs are shown in Figures 6 and 7, where the effect of each disease pathogen can be observed. Although we can see distinctions for each disease, the variation in terms of the number and size of the patches and the unpredictable shapes of the fruits and leaves make the classification task complex.

4.1. Dataset curation and evaluation

To ensure a diverse and representative dataset for training and evaluation, a custom 10-class dataset was collected over a span of six months (July to December 2024) from multiple agricultural fields, experimental farms, and open environments across Karnataka, India. This dataset comprises high-resolution RGB images of fruit and leaf samples infected by bacterial, fungal, and viral pathogens at three distinct infection stages: initial, intermediate, and final, along with healthy samples.

Figure 6
Sample images of different diseases from our fruit datasets. These samples are classified successfully by the proposed method

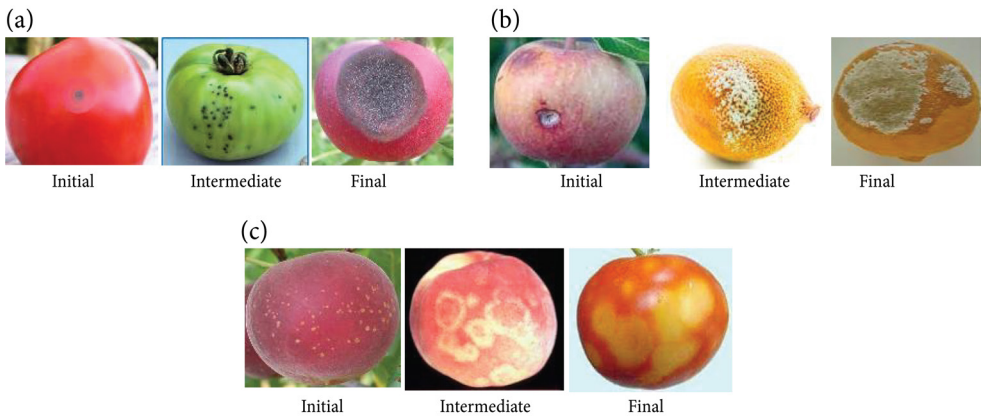


Figure 7
Sample images of different diseases from our leaf datasets. These samples are classified successfully by the proposed method

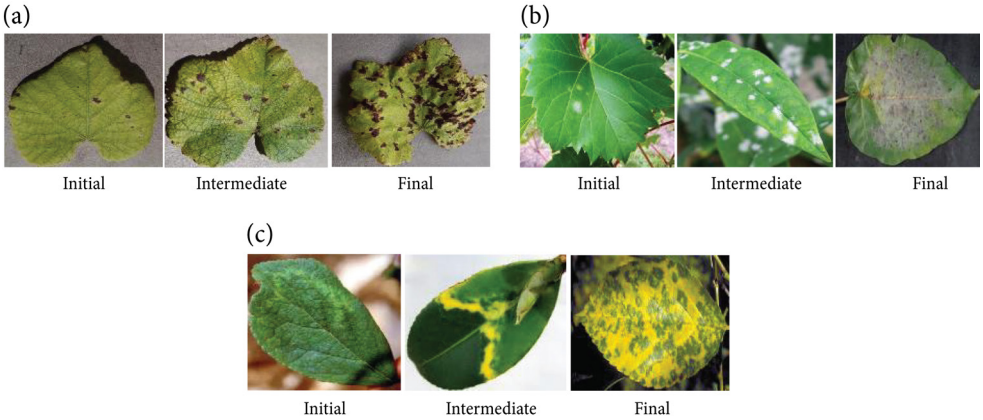


Image Acquisition and Diversity: Images are captured using smartphone cameras and digital SLRs with resolutions ranging from 12 to 24 MP, ensuring both clarity and variability in resolution. The camera-to-subject distance ranged from 0.5 to 1 m, enabling consistent visibility of disease symptoms such as patch color, size, and location. To represent real-world diversity, data is collected under varying conditions, such as natural daylight, shaded conditions, and overcast skies. Similarly, images with different backgrounds, such as plain lab backgrounds and cluttered field environments. Sample of overlapping leaves, stems, soil, noise, blurring, and partial occlusions. This comprehensive capture approach ensures the model is exposed to various real-world variations, enhancing its robustness and generalization ability.

Annotation Protocol: All images are annotated and classified by an expert from the Department of Biotechnology, Davangere University, who has expertise in plant pathology and microbial disease symptoms in horticultural crops. The expert labelled each image into one of ten predefined classes (bacteria, fungus, virus – initial/intermediate/final, and healthy). Visual features such as patch color (yellow for virus, dark for bacteria, white for fungus), shape, and lesion spread were used as key cues for labelling. The labelling process ensured high accuracy and consistency due to the annotator's domain expertise and familiarity with disease progression characteristics.

Dataset Composition: Each of the ten classes includes 500 images, resulting in a balanced dataset of 5000 samples. The dataset includes both fruit and leaf samples, with balanced representation across all disease types and stages. Augmentation techniques such as flipping, color jitter, affine transformations, and rotations further increase sample variability.

NZDLPlantDisease-v1 dataset: This dataset includes images of kiwifruit, apples, pears, avocados, and grapevines from New Zealand agricultural fields. The dataset contains images of multiple diseases on leaves, fruits, and stems under various environmental conditions. In total, there were 1500 images in the bacteria class, 400 images in the fungus class, 1500 images in the healthy class, and 648 images in the virus class [33, 34].

NZDLPlantDisease-v2 dataset: This dataset comprises a diverse collection of images of plant diseases found in New Zealand's vegetables. The bacteria, fungi, viruses, and healthy classes contained 1800, 652, 1800, and 648 images, respectively, yielding a total of 4900 samples. This includes multiple disease categories that affect different vegetables in diverse environmental conditions. Therefore, the high accuracy on these benchmark datasets validates the generalization capability and robustness of the proposed method [33, 34]. These two benchmark datasets are used exclusively for testing purposes. The model is trained and validated solely on the custom-curated dataset from Karnataka, India. This separation ensures that the benchmark test results reflect true out-of-distribution generalization performance. The proposed model is tested on both NZDL datasets without fine-tuning. High classification accuracy on these datasets (ACR > 93%) supports the robustness and adaptability of our model to unseen geographies and crop types.

To demonstrate that the proposed method is superior to existing methods, we implemented four state-of-the-art methods for general image classification [11] and plant disease classification [24]. The methods of Rajalakshmi et al. [35] and Singaravelu and Perumal [36] are developed for banana leaf and plant disease identification, respectively. Zhang and Tu [9] developed a method for the classification of remote sensing images based on Swin Transformer and the Fourier Transform. Tunio et al. [22] proposed a method for plant disease classification based on transformer-fused convolution and Wasserstein domain adaptation. The method was chosen to show that the general image classification method is not effective for the classification of bacteria, viruses, and fungus-infected fruit and leaf images. Similarly, the method was chosen to show that the existing plant disease classification methods may not be robust enough to handle the complex 10-class classification problem.

To evaluate the proposed and existing methods, a confusion matrix was generated, and the Average Classification Rate (ACR) was calculated, which is the mean of the diagonal elements of the confusion matrix.

Implementation Details: For our experiments, we employed the following software and hardware components: Software: OS: Windows 10, Editor: VSCODE 1.97.2, Python: 3.12.9, Optimizer: Adam optimizer, Training Method: Standard supervised learning, Number of Epochs: 30, Hardware: Processor – AMD Ryzen 3200G @3.6 GHz, RAM: 16 GB, HDD: 1 TB. All experiments were conducted using a GPU-enabled environment. The models were trained and tested using a CUDA-compatible device with automatic mixed precision (AMP) to optimize performance.

Hyperparameter and Training Setup: The Tiny Swin Transformer V2 model was implemented using the PyTorch deep learning framework (v2.1.0) and initialized with ImageNet-1K pre-trained weights from the TorchVision model repository. All experiments were conducted on a system equipped with an NVIDIA RTX 3080 GPU (10 GB VRAM), Intel Core i9 CPU, and 16 GB RAM. The input images were resized to 224 × 224 pixels and divided into non-overlapping 4 × 4 patches. A batch size of 16 was used for both training and validation. The model was trained for 10 epochs using the Adam optimizer with an initial learning rate of 3e−4 and a weight decay of 1e−5. A StepLR scheduler was used to reduce the learning rate by a factor of 0.5 every 5 epochs to improve convergence.

To improve generalization, CrossEntropyLoss with label smoothing ($\varepsilon = 0.1$) was used as the loss function. The classification head of the model was modified to include a linear layer that maps the feature vector (768 dimensions) to 256, followed by a ReLU activation, a dropout layer with a rate of 0.5, and a final linear layer outputting logits for 10

classes. Early stopping was employed with a patience of 10 epochs to prevent overfitting. Training was performed using PyTorch's Automatic Mixed Precision (AMP) to accelerate computation and optimize GPU memory usage. Batch normalization and layer normalization are implicitly handled within the Swin Transformer architecture.

4.2. Ablation study

The framework of Swin Transformer V2-Tiny presented in Section 3 comprises several components. To validate the effectiveness and contribution of each component, we conducted experiments on the fruit and leaf datasets and calculated the average classification rate. (i)–(iii) Use of color features: R, G, B, and Gray color spaces. (iv) Use of augmentation to increase the number of samples and diversity to make the model robust. (vi) Use of transfer learning to improve the model performance. (vii) Compared with the baseline Swin Transformer V2-Tiny model to show that the adapted Swin Transformer V2-Tiny is better. (viii) Proposed model without a modified classifier layer. (ix) Proposed model without W-MSA and SW-MSA. (x) Proposed method. The results of all the above experiments are reported in Table 2 for both the fruit and leaf datasets.

Table 2
Average Classification Rate (ACR) of the key steps of the proposed method for fruit and leaf disease classification

SI no.	Key steps	Fruit	Leaf
(i)	R image	79.47	93.29
(ii)	G image	67.21	81.27
(iii)	B image	56.91	89.91
(iv)	Gray image	60.58	83.67
(v)	Without augmentation	90.53	90.87
(vi)	Without transfer learning	13.57	27.79
(vii)	Input images to Swin Transformer V2-Tiny directly	80.32	91.57
(viii)	Without modifying the classifier head of baseline Swin (classifier performance)	69.88	92.92
(xi)	Without W-MSA and SW-MSA	32.74	74.36
(x)	Proposed method	91.04	94.07

As shown in Table 2, all steps contributed to achieving the best results. The proposed method scored the highest average classification rate compared to the individual key steps. This implies that the key components mentioned above were effective. When we compared the performance of different color spaces, the red color space contributed more to both fruit and leaf pathogen classification than other color spaces. This shows that the red color is effective for images of viruses, bacteria, and fungi. A possible reason for this is that the causes or effects of viruses, bacteria, and fungi can be noticed in red spaces compared with other color spaces.

Augmentation techniques, such as flipping, rotation, color jitter, and affine transformations, were used to increase the number of samples and diversity, which helps the model generalize better to unseen data. In addition, by artificially expanding the dataset and simulating real-world conditions, augmentation reduces overfitting and improves the robustness of the model.

Transfer learning is a crucial technique in deep learning that allows models to leverage knowledge from previously learned tasks, leading to faster training, improved performance, and a reduced risk of overfitting. Instead of training a neural network from scratch, a pre-trained model, such as Swin Transformer V2-Tiny trained on ImageNet, is fine-tuned on a new dataset. This approach is especially beneficial when working with limited data, as it enables the model to generalize well by utilizing pre-learned features, such as edges, textures, and shapes. Additionally, transfer learning significantly reduces computational costs by reusing a well-trained feature extractor and modifying only the final layers to fit the specific classification task. In our work, the Swin Transformer V2-Tiny implementation, transfer learning was applied by loading a pre-trained Swin Transformer V2-Tiny model and replacing its final classification layer to adapt to the 10-class problem, ensuring that the model retained useful features while specializing in the dataset.

When input images were supplied directly to Swin Transformer V2-Tiny for classification, the results were not as high as those of the proposed method. Therefore, Swin Transformer V2-Tiny alone is insufficient for achieving high results. Similarly, if we test the proposed model without the modified classifier layer, the results are not as high as those of the proposed method. This indicates that the proposed modified classifier layer contributes to achieving high results. In summary, the above analysis showed that all the main steps mentioned in this study are effective and useful for obtaining high accuracy in the classification of fruit and leaf diseases at different levels.

Table 3
Confusion matrix and ACR (in %) of the proposed method on fruit disease classification

Fruit class	B-F	B-Ini	B-Int	F-Ini	F-F	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	93.47	0.41	0.41	5.30	0.41	0	0	0	0	0
B-Ini	5.77	84.61	3.85	0	3.85	0	0	0	1.92	0
B-Int	1.92	0	90.38	3.86	0	0	1.92	0	1.92	0
F-Ini	9.09	3.03	0	84.85	0	3.03	0	0	0	0
F-F	0	0	0	0	98.04	1.96	0	0	0	0
F-Int	1.88	0	0	0	7.55	90.57	0	0	0	0
Healthy	0	0	0	0	0	0	100	0	0	0
V-F	0	0	0	0	0	0	0	100	0	0
V-Ini	0	0	0	0	0	0	0	0	86.67	13.33
V-Int	0	0	0	0	0	0	0	18.18	0	81.82
					ACR = 91.04					

Note: B-F: Bacteria final stage, B-Ini: Bacteria initial stage, B-Int: Bacteria intermediate stage, F-Ini: Fungus initial stage, F-F: Fungus final tsage, F-Int: Fungus intermediate stage, Healthy class, V-F: Virus final stage, V-Ini: Virus initial stage, and V-Int: Virus intermediate stage.

4.3. Experiments on fruit and leaf disease classification

The confusion matrix and average classification rates of the proposed and existing methods — SwinFR Tables 5 & 8 [9]; TFC-WDA Tables 4 & 7 [22]; NDCNN Tables 9–10 [35]; and DCoS-WOR-SNN Tables 11–12 [36]— for the fruit and leaf datasets are presented in Tables 3–12. As shown in Tables 3–12, the proposed method outperforms the existing methods for both the fruit and leaf datasets in terms of the average classification rate. The existing methods report

Table 4
Confusion matrix and ACR (in %) on fruit disease classification using TFC-WDA

Fruit class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	96.92	3.08	0	0	0	0	0	0	0	0
B-Ini	47.47	47.47	0	0	0	5.26	0	0	0	0
B-Int	95.45	4.55	0	0	0	0	0	0	0	0
F-Ini	48.15	22.22	0	0	0	29.63	0	0	0	0
F-F	3.51	0	0	0	89.47	7.02	0	0	0	0
F-Int	6.06	9.09	0	0	0	84.85	0	0	0	0
Healthy	0	0	0	0	0	0	100	0	0	0
V-F	42.86	14.29	0	0	0	35.71	7.14	0	0	0
V-Ini	6.25	25.00	0	0	0	62.50	6.25	0	0	0
V-Int	5.55	16.67	0	0	0	72.22	5.55	0	0	0

ACR = 41.86

Table 5
Confusion matrix and ACR (in %) on fruit disease classification using SwinFR

Fruit class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	82.13	5.96	0	0	9.72	0.31	1.88	0	0	0
B-Ini	18.42	69.74	0	0	3.95	0	7.89	0	0	0
B-Int	40.84	23.94	7.05	0	18.31	5.63	4.23	0	0	0
F-Ini	40.00	37.5	7.5	0	7.5	0	7.5	0	0	0
F-F	40.32	0.77	0	0	37.98	8.53	12.40	0	0	0
F-Int	24.61	6.15	1.54	0	32.31	23.08	12.31	0	0	0
Healthy	31.82	2.27	2.27	0	2.27	0	61.37	0	0	0
V-F	50.00	14.28	0	0	14.29	14.29	7.14	0	0	0
V-Ini	31.25	12.5	0	0	25.00	0	31.25	0	0	0
V-Int	33.33	11.11	5.56	0	11.11	0	38.89	0	0	0

ACR = 51.89

Table 6
Confusion matrix and ACR (in %) of the proposed method on leaf disease classification

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	98.25	0	1.75	0	0	0	0	0	0	0
B-Ini	0	95.42	4.58	0	0	0	0	0	0	0
B-Int	0	1.40	98.60	0	0	0	0	0	0	0
F-F	0	0	0	93.33	2.22	4.45	0	0	0	0
F-Ini	0	0	0	0	81.40	11.63	0	0	6.97	0
F-Int	0	0	0	8.33	2.78	88.89	0	0	0	0
Healthy	0	0	0	0	0	0	100	0	0	0
V-F	2.27	0	0		0	0	0	97.73	0	0
V-Ini	0	0	0	0	5.71	0	0	0	91.43	2.85
V-Int	0	0	0	0	0	0	0	4.35	0	95.65

ACR = 94.07

Table 7
Confusion matrix and ACR (in %) on fruit disease classification using TFC-WDA

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	94.74	0	0	2.63	0	0	0	2.63	0	0
B-Ini	0	100	0	0	0	0	0	0	0	0
B-Int	0.86	0	99.14	0	0	0	0	0	0	0
F-Ini	0	0	0	100	0	0	0	0	0	0
F-F	0	0	0	0	45.26	8.42	0	0	0	0
F-Int	0	0	1.06	38.30	8.51	52.13	0	0	0	0
Healthy	0	0	0	0	0	0	100	0	0	0
V-F	0	0	0	0	0	0	16.00	82.00	2.00	0
V-Ini	0	3.23	0	0	0	0	6.45	0	90.32	0
V-Int	0	0	0	0	0	0	94.74	5.26	0	0
ACR = 76.36										

Table 8
Confusion matrix and ACR (in %) on fruit disease classification using SwinFR

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	4.11	0	93.83	1.03	0	0	0	1.03	0	0
B-Ini	0	40.12	58.45	0.14	0.72	0.29	0	0.14	0.14	0
B-Int	0	4.33	94.89	0.39	0	0	0	0.33	0	0.06
F-Ini	0	0	2.21	87.62	0	3.54	0	5.75	0	0.88
F-F	0	26.15	28.72	8.20	4.62	19.49	0	5.64	6.67	0.51
F-Int	0	5.31	24.15	19.81	0	37.68	0.97	10.63	0.48	0.97
Healthy	0	0	1.22	0	0	0	97.56	1.22	0	0
V-F	0	1.08	21.40	4.81	0.53	0.53	2.14	68.98	0	0.53
V-Ini	0	0.65	24.18	0.65	1.32	0.65	0	15.69	56.21	0.65
V-Int	0	0	6.19	4.12	0	0	9.28	0	0	80.41
ACR = 57.22										

Table 9
Confusion matrix and ACR (in %) on fruit disease classification using NDCNN

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	84.62	0	0	15.38	0	0	0	0	0	0
B-Ini	52.63	0	0	47.37	0	0	0	0	0	0
B-Int	34.09	0	0	65.91	0	0	0	0	0	0
F-Ini	55.56	0	0	44.44	0	0	0	0	0	0
F-F	21.05	0	0	78.95	0	0	0	0	0	0
F-Int	15.62	0	0	84.38	0	0	0	0	0	0
Healthy	94.44	0	0	5.56	0	0	0	0	0	0
V-F	35.71	0	0	64.29	0	0	0	0	0	0
V-Ini	18.75	0	0	81.25	0	0	0	0	0	0
V-Int	22.22	0	0	77.78	0	0	0	0	0	0
ACR = 12.91										

Table 10
Confusion matrix and ACR (in %) on fruit disease classification using NDCNN

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	68.42	0	13.16	2.63	0	0	0	10.53	5.26	0
B-Ini	0	92.59	4.63	0	1.85	0	0	0	0.93	0
B-Int	1.29	1.72	97.00	0	0	0	0	0	0	0
F-Ini	0	0.97	0	98.06	0	0	0	0.97	0	0
F-F	0	4.21	0	16.84	49.47	26.32	0	0	3.16	0
F-Int	0	0	0	59.14	7.53	32.26	0	0	0	1.08
Healthy	0	5.56	0	0	16.67	8.33	63.89	0	2.78	2.78
V-F	2.00	0	0	8.00	6.00	0	8.00	60.00	16.00	0
V-Ini	0	0	0	0	0	0	0	0	96.67	3.33
V-Int	0	0	16.67	5.56	0	0	16.67	5.56	11.11	44.44

ACR = 70.28

Table 11
Confusion matrix and ACR (in %) on fruit disease classification using DCoS-WOR-SNN

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	0	0	100.0	0	0	0	0	0	0	0
B-Ini	0	0	100.0	0	0	0	0	0	0	0
B-Int	0	0	100.0	0	0	0	0	0	0	0
F-Ini	0	0	100.0	0	0	0	0	0	0	0
F-F	0	0	100.0	0	0	0	0	0	0	0
F-Int	0	0	100.0	0	0	0	0	0	0	0
Healthy	0	0	100.0	0	0	0	0	0	0	0
V-F	0	0	100.0	0	0	0	0	0	0	0
V-Ini	0	0	100.0	0	0	0	0	0	0	0
V-Int	0	0	100.0	0	0	0	0	0	0	0

ACR = 10.00

Table 12
Confusion matrix and ACR (in %) on fruit disease classification using DCoS-WOR-SNN

Leaf class	B-F	B-Ini	B-Int	F-F	F-Ini	F-Int	Healthy	V-F	V-Ini	V-Int
B-F	100.0	0	0	0	0	0	0	0	0	0
B-Ini	100.0	0	0	0	0	0	0	0	0	0
B-Int	100.0	0	0	0	0	0	0	0	0	0
F-Ini	100.0	0	0	0	0	0	0	0	0	0
F-F	100.0	0	0	0	0	0	0	0	0	0
F-Int	100.0	0	0	0	0	0	0	0	0	0
Healthy	100.0	0	0	0	0	0	0	0	0	0
V-F	100.0	0	0	0	0	0	0	0	0	0
V-Ini	100.0	0	0	0	0	0	0	0	0	0
V-Int	100.0	0	0	0	0	0	0	0	0	0

ACR = 10.00

Table 13
Precision, recall, F1-score, and per class accuracy of the proposed method on fruit and leaf disease classification

	Fruit dataset				Leaf dataset			
Fruit class	**Precision**	**Recall**	**F1-score**	**Per class accuracy**	**Precision**	**Recall**	**F1-score**	**Per class accuracy**
B-F	0.84	0.93	0.88	93.47	0.98	0.98	0.98	98.25
B-Ini	0.95	0.85	0.90	84.61	0.99	0.95	0.97	95.42
B-Int	0.95	0.90	0.92	90.38	0.93	0.99	0.96	98.60
F-F	0.91	0.85	0.88	84.85	0.91	0.93	0.92	93.33
F-Ini	0.92	0.98	0.95	98.04	0.87	0.81	0.84	81.40
F-Int	0.96	0.91	0.93	90.57	0.91	0.89	0.90	88.89
Healthy	0.98	1.00	0.99	100	1.00	1.00	1.00	100
V-F	0.85	1.00	0.92	100	0.96	0.98	0.97	97.73
V-Ini	0.87	0.87	0.87	86.67	0.93	0.91	0.92	91.43
V-Int	0.86	0.82	0.84	81.82	0.97	0.96	0.96	95.65

Table 14
Precision, recall, F1-score, and per class accuracy on fruit and leaf disease classification

	Fruit dataset				Leaf dataset			
Fruit class	**Precision**	**Recall**	**F1-score**	**Per class accuracy**	**Precision**	**Recall**	**F1-score**	**Per class accuracy**
B-F	0.24	0.97	0.39	96.92	0.95	0.95	0.95	94.74
B-Ini	0.27	0.47	0.34	47.47	0.93	1.00	0.96	100
B-Int	0	0	0	0	0.99	0.99	0.99	99.14
F-F	0	0	0	0	0.71	1.00	0.83	100
F-Ini	0.80	0.89	0.84	89.47	0.79	0.45	0.57	45.26
F-Int	0.33	0.85	0.48	84.85	0.73	0.52	0.61	52.13
Healthy	0.93	1.00	0.97	100	0.94	1.00	0.97	100
V-F	0	0	0	0	0.89	0.82	0.85	82.00
V-Ini	0	0	0	0	0.95	0.90	0.92	90.32
V-Int	0	0	0	0	1.00	0	0	0

Table 15
Precision, recall, F1-score, and per class accuracy on fruit and leaf disease classification

	Fruit dataset				Leaf dataset			
Fruit class	**Precision**	**Recall**	**F1-score**	**Per class accuracy**	**Precision**	**Recall**	**F1-score**	**Per class accuracy**
B-F	0.24	0.82	0.37	82.13	0.85	0.94	0.89	93.83
B-Ini	0.40	0.70	0.51	69.47	0.38	0.40	0.39	40.12
B-Int	0.35	0.07	0.12	7.05	0.37	0.95	0.53	94.89
F-F	0.19	0.08	0.11	7.50	0.47	0.88	0.61	87.62
F-Ini	0.28	0.38	0.32	37.98	0.32	0.05	0.09	4.62
F-Int	0.34	0.23	0.27	23.08	0.36	0.38	0.37	37.68
Healthy	0.39	0.61	0.48	61.37	0.92	0.98	0.95	97.56
V-F	0	0	0	0	0.55	0.69	0.61	68.98
V-Ini	0	0	0	0	0.82	0.56	0.66	56.21
V-Int	0	0	0	0	0.60	0.80	0.88	80.41

Table 16
Precision, recall, F1-score, and per class accuracy on fruit and leaf disease classification

	Fruit dataset				Leaf dataset			
Fruit class	**Precision**	**Recall**	**F1-score**	**Per class accuracy**	**Precision**	**Recall**	**F1-score**	**Per class accuracy**
B-F	0.39	0.85	0.53	84.62	0.87	0.68	0.76	68.42
B-Ini	0	0	0	0	0.90	0.93	0.91	92.59
B-Int	0	0	0	0	0.95	0.97	0.96	97.00
F-F	0.07	0.44	0.12	78.95	0.57	0.98	0.72	98.06
F-Ini	0	0	0	0	0.72	0.49	0.59	49.47
F-Int	0	0	0	0	0.52	0.32	0.40	32.26
Healthy	0	0	0	0	0.77	0.64	0.70	63.89
V-F	0	0	0	0	0.83	0.60	0.70	60.00
V-Ini	0	0	0	0	0.63	0.97	0.76	96.67
V-Int	0	0	0	0	0.73	0.44	0.55	44.44

Table 17
Precision, recall, F1-score, and per class accuracy on fruit and leaf disease classification

	Fruit dataset				Leaf dataset			
Fruit class	**Precision**	**Recall**	**F1-score**	**Per class accuracy**	**Precision**	**Recall**	**F1-score**	**Per class accuracy**
B-F	0.042	0.100	0.059	0.100	0	0	0	0
B-Ini	0	0	0	0	0.046	0.100	0.063	0.100
B-Int	0	0	0	0	0	0	0	0
F-F	0	0	0	0	0	0	0	0
F-Ini	0	0	0	0	0	0	0	0
F-Int	0	0	0	0	0	0	0	0
Healthy	0	0	0	0	0	0	0	0
V-F	0	0	0	0	0	0	0	0
V-Ini	0	0	0	0	0	0	0	0
V-Int	0	0	0	0	0	0	0	0

poor results because Tunio et al.'s method [22] is good for plant but not fruit images, while Zhang and Tu's method [9] is good for general image classification. Because the scope of the existing methods is limited to particular datasets and cases, they do not perform well for the 10-class classification of fruit and leaf datasets. It is evident from the performance of the existing methods on fruit and leaf datasets that Tunio et al.'s method [22] performed better for the leaf dataset and worse for the fruit datasets than Zhang and Tu's [9]. In addition, Zhang and Tu [9] and Singaravelu and Perumal [36] obtained almost the same results for the fruit and leaf datasets. This is because the method considers fruit and leaf images as general images.

However, as discussed in the ablation study experiments, the key steps proposed in this study are effective. The combination of Color spaces, Transfer Learning, and Swin Transformer V2-Tiny enhances the generalization ability, and the proposed method is superior in terms of the average classification rate compared with existing methods on both fruit and leaf datasets. Furthermore, the proposed method achieved similar results for both datasets. This justifies that the proposed method is consistent and domain-independent.

To validate the above statement, we also calculated precision, recall, F1-score, and accuracy per class for the proposed and existing methods on fruit and leaf dataset as reported in Tables 13–17 (Table 14 [22]; Table 15 [9]; Table 16 [35]; Table 17 [36]). It is observed from Tables 13–17 that the proposed method outperforms all the existing methods in terms of precision, recall, F1-score, and accuracy per class

Table 18
ACR of the proposed and existing methods on NZDLPlantDisease-v1, NZDLPlantDisease-v2 (in %)

Methods	**NZDLPlant-Disease-v1**	**NZDLPlant-Disease-v2**
Proposed method	95.44	93.64
Tunio et al. [22]	52.60	52.15
Zhang et al. [11]	36.92	50.53
Rajalakshmi et al. [35]	17.09	7.00
Singaravelu and Perumal [36]	7.14	7.69

Table 19
The average classification rate of the proposed and existing methods on scaled, rotated, and distorted images

	Random scaling up and down		Random rotations		Different levels of Gaussian noise and blur	
Methods	**Fruit**	**Leaf**	**Fruit**	**Leaf**	**Fruit**	**Leaf**
Proposed method	81.88	87.51	80.44	94.80	69.24	73.04
Tunio et al. [22]	46.86	70.50	37.47	63.70	30.24	59.56
Zhang et al. [9]	47.85	61.29	48.61	63.22	49.49	61.64
Rajalakshmi et al. [35]	10.84	65.81	10.97	74.59	12.51	77.64
Singaravelu and Perumal [36]	10.00	10.00	10.00	10.00	10.00	10.00

on both fruit and leaf datasets. This shows that the proposed method is domain-independent and generic. However, since the existing methods are developed with specific objectives, the performance of the existing methods is inferior to the proposed method on both fruit and leaf datasets. In addition, the existing methods are not capable of handling 10 classes of diseases.

4.4. Experiments for classification on two benchmark datasets

The performance of the proposed and existing methods was tested on two benchmark datasets, as reported in Table 18, where it can be seen that the proposed method is the best for both datasets compared to the existing methods. Although the datasets were more complex than our fruit and leaf datasets, the proposed method achieved consistent results for both datasets. The reasons for the successful classification of the proposed method and the poor results of the existing methods are the same as those stated in the previous section. Overall, when we analyzed the experiments on our and benchmark datasets, the proposed method was effective, consistent, and generic.

4.5. Experiments on robustness analysis

To demonstrate that the proposed method is robust to distortion, rotation, noise, blur, and scaling, the average classification rate was calculated for different experiments on our dataset, and the results are reported in Table 19. The results listed in Table 19 show that the proposed method obtained almost consistent results for different scaling, rotation, blur, and noise compared with the existing methods. This indicates that the method is invariant to rotation, scaling, noise, and blur. This includes that the extracted features are insensitive to noise, blur, and the effects of rotation and scaling. However, the existing method lacks a generic nature and robustness, and the methods are inferior to the proposed method in various experiments.

Challenges: While the proposed Tiny Swin Transformer V2 model demonstrates strong performance in controlled experimental settings, deploying it in real-world agricultural environments presents several challenges. One major limitation is the computational demand of transformer-based models, which can be unsuitable for real-time inference on low-power edge devices such as mobile phones or drones. Although Tiny Swin V2 is relatively lightweight compared to larger ViTs, further optimization (e.g., quantization, pruning, or model distillation) may be required to ensure compatibility with embedded systems.

Additionally, the variability in image acquisition conditions—such as inconsistent lighting, background clutter, occlusion from other leaves or fruits, camera motion blur, or weather effects—can impact model accuracy. While data augmentation helps simulate these variations, unpredictable field conditions may still affect robustness. There is also a need for cross-geographic validation, as disease symptoms can vary between regions, crop varieties, and climate zones. Finally, ethical deployment must ensure that such automated systems are used to assist rather than replace agronomists and plant health experts. Providing confidence scores, explainable AI outputs, and human-in-the-loop verification is essential for responsible adoption in agricultural decision-making pipelines.

5. Conclusion and Future Work

In this study, we adapted the Tiny Swin Transformer V2 model for the classification of fruit and leaf images infected by viruses, bacteria, and fungi at different levels. As stated in Section 3, visual features such as yellow, dark, and white patches are important cues for the classification of fruit and leaf images infected by viruses, bacteria, and fungi. Inspired by vision transformers that extract visual features accurately, the proposed work adapts the Tiny Swin Transformer V2 for the classification of three stages of viruses, bacteria, and fungi on fruits and leaves. Unlike the baseline Swin Transformer and Swin Transformer V2-Tiny, the proposed study chose the Tiny Swin Transformer V2 for successful classification. This is because the proposed method is more efficient, accurate, and adaptable to different situations than the baseline models. The results on the fruit and leaf dataset and two benchmark datasets, and a comparative study with the existing methods, show that the proposed method is effective, robust, domain-independent, and consistent. However, the proposed method did not perform well when it was trained on samples from other fruit and leaf datasets. This shows that our method is sensitive to the training samples. This can be solved by proposing a combination of GANS, transformers, and diffusion models. This is beyond the scope of the present work and can be extended to the near future. Since the scope of the work is to address the challenge of fruit and leaf disease identification using an adaptive Swin transformer, the present work does not focus on theory to justify the hypothesis of the proposed method. Therefore, this is beyond the scope of the work, and hence it can be considered future work.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Shivakumara Palaiahnakote is the Editor-in-Chief for *Artificial Intelligence and Applications*, and he was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Poornima Basatti Hanuma Gowda: Software, Data curation, Writing – original draft, Visualization. **Basavanna Mahadevappa:** Formal analysis, Investigation, Supervision, Project administration. **Shivakumara Palaiahnakote:** Conceptualization, Methodology. **Muhammad Hammad Saleem:** Validation, Writing – review & editing. **Niranjan Mallappa Hanumanthu:** Resources.

References

[1] Saraswat, S., Singh, P., Kumar, M., & Agarwal, J. (2024). Advanced detection of fungi-bacterial diseases in plants using modified deep neural network and DSURF. *Multimedia Tools and Applications*, *83*(6), 16711–16733. https://doi.org/10.1007/s11042-023-16281-1

[2] Gaikwad, S. S., Rumma, S. S., & Hangarge, M. (2021). Identification of fungi infected leaf diseases using deep learning techniques. *Turkish Journal of Computer and Mathematics Education*, *12*(6), 5618–5625.

[3] Gaikwad, S. S., Rumma, S. S., & Hangarge, M. (2022). Fungi affected fruit leaf disease classification using deep CNN architecture. *International Journal of Information Technology*, *14*(7), 3815–3824. https://doi.org/10.1007/s41870-022-00860-w

[4] Li, B., Zhang, J., Cao, J., Zhang, J., & Gao, L. (2023). PMST: A parallel and miniature Swin transformer for logo detection. *Digital Signal Processing*, *140*, 104102. https://doi.org/10.1016/j.dsp.2023.104102

[5] Hu, Z., Wang, Z., Jin, Y., & Hou, W. (2023). VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. *Computer Methods and Programs in Biomedicine*, *229*, 107291. https://doi.org/10.1016/j.cmpb.2022.107291

[6] Zeng, C., Kwong, S., & Ip, H. (2023). Dual Swin-transformer based mutual interactive network for RGB-D salient object detection. *Neurocomputing*, *559*, 126779. https://doi.org/10.1016/j.neucom.2023.126779

[7] He, W., Ren, J., Bai, R., & Jiang, X. (2025). Radar gait recognition using dual-branch Swin Transformer with asymmetric attention fusion. *Pattern Recognition*, *159*, 111101. https://doi.org/10.1016/j.patcog.2024.111101

[8] Hu, J., Pang, T., Peng, B., Shi, Y., & Li, T. (2025). A small object detection model for drone images based on multi-attention fusion network. *Image and Vision Computing*, *155*, 105436. https://doi.org/10.1016/j.imavis.2025.105436

[9] Zhang, J., & Tu, Y. (2025). SwinFR: Combining SwinIR and fast Fourier for super-resolution reconstruction of remote sensing images. *Digital Signal Processing*, *159*, 105026. https://doi.org/10.1016/j.dsp.2025.105026

[10] Zhou, J., Yang, D., Song, T., Ye, Y., Zhang, X., & Song, Y. (2024). Improved YOLOv7 models based on modulated deformable convolution and Swin transformer for object detection in fisheye images. *Image and Vision Computing*, *144*, 104966. https://doi.org/10.1016/j.imavis.2024.104966

[11] Pal, U., Halder, A., Shivakumara, P., & Blumenstein, M. (2024). A comprehensive review on text detection and recognition in scene images. *Artificial Intelligence and Applications*, *2*(4), 229–249. https://doi.org/10.47852/bonviewAIA42022755

[12] Kang, M., Zhao, J., & Farid, F. (2024). Implications of classification models for patients with chronic obstructive pulmonary disease. *Artificial Intelligence and Applications*, *2*(2), 97–106. https://doi.org/10.47852/bonviewAIA32021406

[13] Gupta, S., & Tripathi, A. K. (2024). Fruit and vegetable disease detection and classification: Recent trends, challenges, and future opportunities. *Engineering Applications of Artificial Intelligence*, *133*, 108260. https://doi.org/10.1016/j.engappai.2024.108260

[14] Laim, Biang & Y, Laim & Kumar, Rahul. (2023). Fruit disease classification through machine learning. *Concurrency and Computation Practice and Experience*. *5*, 74–81. https://www.researchgate.net/publication/371416779

[15] Aboelenin, S., Elbasheer, F. A., Eltoukhy, M. M., El-Hady, W. M., & Hosny, K. M. (2025). A hybrid framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer. *Complex & Intelligent Systems*, *11*(2), 1–17. https://doi.org/10.1007/s40747-024-01764-x

[16] Xu, X., Yang, G., Wang, Y., Shang, Y., Hua, Z., Wang, Z., & Song, H. (2024). Plant leaf disease identification by parameter-efficient transformer with adapter. *Engineering Applications of Artificial Intelligence*, *138*, 109466. https://doi.org/10.1016/j.engappai.2024.109466

[17] Megalingam, R. K., Menon, G. G., Binoj, S., Sai, D. A., Kunnambath, A. R., & Manoharan, S. K. (2024). Cowpea leaf disease identification using deep learning. *Smart Agricultural Technology*, *9*, 100662. https://doi.org/10.1016/j.atech.2024.100662

[18] Das, A., Pathan, F., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2025). Deep learning-based classification, detection, and segmentation of tomato leaf diseases: A state-of-the-art review. *Artificial Intelligence in Agriculture*, *15*(2), 192–220. https://doi.org/10.1016/j.aiia.2025.02.006

[19] Wu, P., Liu, J., Jiang, M., Zhang, L., Ding, S., & Zhang, K. (2025). Tea leaf disease recognition using attention convolutional neural network and handcrafted features. *Crop Protection*, *190*, 107118. https://doi.org/10.1016/j.cropro.2025.107118

[20] Barman, U., Sarma, P., Rahman, M., Deka, V., Lahkar, S., Sharma, V., & Saikia, M. J. (2024). Vit-SmartAgri: Vision transformer and smartphone-based plant disease detection for smart agriculture. *Agronomy*, *14*(2), 327. https://doi.org/10.3390/agronomy14020327

[21] Hemalatha, S., & Jayachandran, J. J. B. (2024). A multitask learning-based vision transformer for plant disease localization and classification. *International Journal of Computational Intelligence Systems*, *17*(1), 188. https://doi.org/10.1007/s44196-024-00597-3

[22] Tunio, M. H., ping Li, J., Zeng, X., Ahmed, A., Shah, S. A., Shaikh, H. U., ... & Yahya, I. A. (2024). Advancing plant disease classification: A robust and generalized approach with transformer-fused convolution and Wasserstein domain adaptation. *Computers and Electronics in Agriculture*, *227*, 109574. https://doi.org/10.1016/j.compag.2024.109574

[23] Singh, A. K., Rao, A., Chattopadhyay, P., Maurya, R., & Singh, L. (2024). Effective plant disease diagnosis using Vision Transformer trained with leafy-generative adversarial network-generated images. *Expert Systems with Applications*, *254*, 124387. https://doi.org/10.1016/j.eswa.2024.124387

[24] Liu, W., & Zhang, A. (2025). Plant disease detection algorithm based on efficient Swin transformer. *Computers, Materials & Continua*, *82*(2). https://doi.org/10.32604/cmc.2024.058640

[25] Subramanian, K., Hajamohideen, F., Viswan, V., Shaffi, N., & Mahmud, M. (2024). Exploring intervention techniques for Alzheimer's disease: Conventional methods and the role of AI in advancing care. *Artificial Intelligence and Applications*, *2*(2), 59–77. https://doi.org/10.47852/bonviewAIA42022497

[26] Balasundaram, A., Sundaresan, P., Bhavsar, A., Mattu, M., Kavitha, M. S., & Shaik, A. (2025). Tea leaf disease detection using segment anything model and deep convolutional neural networks. *Results in Engineering*, *25*, 103784. https://doi.org/10.1016/j.rineng.2024.103784

[27] Hasan, M., Gani, R., Rashid, M. R. A., Isty, M. N., Kamara, R., & Tarin, T. K. (2025). Smartphone image dataset for radish plant leaf disease classification from Bangladesh. *Data in Brief*, *58*, 111263. https://doi.org/10.1016/j.dib.2024.111263

[28] Ni, Q., Zuo, Y., Zhi, Z., Shi, Y., Liu, G., & Ou, Q. (2024). Diagnosis of corn leaf diseases by FTIR spectroscopy combined with machine learning. *Vibrational Spectroscopy*, *135*, 103744. https://doi.org/10.1016/j.vibspec.2024.103744

[29] Petchiammal, A., & Murugan, D. (2025). Automated paddy leaf disease identification using visual leaf images based on nine pre-trained models approach. *Procedia Computer Science*, *252*, 118–126. https://doi.org/10.1016/j.procs.2024.12.013

[30] Akhter, Y., & Saxena, S. (2025). Association of Catharanthus yellow mosaic virus and a novel Tomato leaf curl Lucknow betasatellite with papaya leaf curl disease (PaLCD) in India. *The Microbe*, *6*, 100256. https://doi.org/10.1016/j.microb.2025.100256

[31] Pakruddin, B., & Hemavathy, R. (2024). A comprehensive standardized dataset of numerous pomegranate fruit diseases for deep learning. *Data in Brief*, *54*, 110284. https://doi.org/10.1016/j.dib.2024.110284

[32] Siripatrawan, U., & Makino, Y. (2024). Hyperspectral imaging coupled with machine learning for classification of anthracnose infection on mango fruit. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *309*, 123825. https://doi.org/10.1016/j.saa.2023.123825

[33] Saleem, M. H., Potgieter, J., & Arif, K. M. (2022a). A performance-optimized 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

[34] Saleem, M. H., Potgieter, J., & Arif, K. M. (2022b). A weight optimization-based transfer learning approach for plant disease detection of New Zealand vegetables. *Frontiers in Plant Science*, *13*, 1008079. https://doi.org/10.3389/fpls.2022.1008079

[35] Rajalakshmi, N. R., Saravanan, S., Arunpandian, J., Mathivanan, S. K., Jayagopal, P., Mallik, S., & Qin, G. (2025). Early detection of banana leaf disease using novel deep convolutional neural network. *Journal of Data Science and Intelligent Systems*, *3*(3), 192–199. https://doi.org/10.47852/bonviewJDSIS42021530

[36] Singaravelu, P., & Perumal, E. (2025). Innovative solutions for plant disease identification: Leveraging DCoS-WOR and spiking neural networks. *Expert Systems with Applications*, *282*, 127399. https://doi.org/10.1016/j.eswa.2025.127399

How to Cite: Gowda, P. B. H., Mahadevappa, B., Palaiahnakote, S., Saleem, M. H., & Hanumanthu, N. M. (2025). Adaptive Swin Transformer V2-Tiny Based Model for Classification of Bacteria, Fungus, Virus, and Healthy Fruit and Leaf Images. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52026081