



A dual track deep fusion network for citrus disease classification using group shuffle depthwise feature pyramid and Swin transformer

R. Karthik ^{a,*}, Sameeha Hussain ^b, Timothy Thomas George ^c, Rashmi Mishra ^b

^a Centre for Cyber Physical Systems (CCPS), Vellore Institute of Technology, Chennai, India

^b School of Electronics Engineering, Vellore Institute of Technology, Chennai, India

^c School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India



ARTICLE INFO

Keywords:

Deep learning
Citrus plant disease
Dual path network
Swin transformer
Shuffle attention

ABSTRACT

Citrus fruits are globally significant crops, valued for their unique taste, nutritional value, and versatility in culinary applications. However, these crops are vulnerable to diseases that can substantially reduce their yield and quality, leading to significant financial losses. It is crucial to efficiently identify and manage diseases affecting citrus fruits to ensure food security and support sustainable agriculture. Traditionally, identifying symptoms associated with these diseases requires expert scientific and observational skills. With the advent of deep learning methods, it has become possible to automatically identify disease patterns from images of affected plants. In this research, a novel dual branch deep-learning network was proposed for the classification of citrus plant diseases. The Group Shuffle Depthwise Feature Pyramid (GSDFP), which constitutes the first branch, utilizes convolution blocks to extract local features. The second branch leverages the Swin transformer to integrate global contextual awareness into the extracted features and facilitates the learning process through long-term dependencies. Subsequently, the features from the two branches are fused and passed through a shuffle attention module, effectively capturing contextual relationships between them. The effectiveness of the proposed approach was validated on two benchmark datasets, namely the Citrus Plant Dataset and the Citrus Disease Image Gallery Dataset. The proposed network obtained a classification accuracy of 98.19%.

1. Introduction

Citrus fruits are a major global commodity, with a significant share of the world trade in horticultural products (Matheyambath et al., 2016). The export of citrus fruits plays a significant role in the economies of several countries, including Brazil, the United States, and Spain. Collectively, these nations produce 34 million tons of citrus fruits annually (Zhong and Nicolosi, 2020). Citrus fruits are a versatile ingredient in the food industry and are used in various preparations such as juices, jellies, and baked goods. Moreover, they contain a range of nutrients that provide relief from gastrointestinal issues and are believed to possess preventive properties for diseases like diabetes, cancer, and neurological disorders (Saini et al., 2022). The use of citrus fruits extends beyond the food industry, with citrus extracts being widely used in health and wellness products. Nevertheless, citrus cultivation is challenging due to various factors such as their high sensitivity to cold weather, soil acidity requirements, and susceptibility to a variety of crop

diseases. Five common diseases that pose a threat to citrus production are Citrus Greening, Anthracnose, Citrus Black Spot (CBS), Citrus Canker (CC), and Sweet Orange Scab (Ferreira et al., 2021).

Citrus cultivators face a major challenge in controlling diseases affecting their crops. The excessive use of fungicides and other chemicals has raised environmental concerns, and the associated costs have become a burden for farmers, leading to significant financial losses (Sampathkumar and Rajeswari, 2020a). To address this issue, various alternatives are being explored to eliminate the use of harmful chemicals. Currently, farmers rely on manual disease diagnosis, which involves consulting with agricultural specialists to visually inspect citrus plants. However, this approach is time-consuming, labor-intensive, and impractical for early disease identification. Therefore, there is a growing awareness of the need to adopt the latest technologies for commercial citrus cultivation, as manual disease detection techniques have limitations (Prabu and Chelliah, 2022).

This paves the way to employ Computer-Aided Diagnosis (CAD)-

* Corresponding author.

E-mail addresses: r.karthik@vit.ac.in (R. Karthik), sameeha.hussain2019@vitstudent.ac.in (S. Hussain), timothythomas.george2020@vitstudent.ac.in (T.T. George), rashmi.mishra2019@vitstudent.ac.in (R. Mishra).

based systems in the field of agriculture to aid in effective diagnoses. Computer-aided diagnosis is an emerging field that utilizes digital imaging and computer vision techniques to facilitate the accurate and efficient diagnosis of diseases (Sampathkumar and Rajeswari, 2020b). This approach has gained momentum due to the rapid advancements in image processing and Machine Learning (ML) techniques, which have enhanced image classification techniques. However, traditional ML algorithms require manual feature extraction for disease detection, which can be time-consuming and resource-intensive (Appiah et al., 2019; Bashar, 2019; Wei et al., 2019). In contrast, Deep Learning (DL) approaches have gained popularity as they enable automatic feature extraction, leading to higher accuracy in disease detection. This is due to the fact that DL attempts to extract high-level characteristics directly from data, which reduces the time and effort required to construct a feature extractor for each problem (Sarker, 2021). Although training a DL algorithm may take a longer duration due to a large number of parameters in the model, DL algorithms require minimal time for testing as compared to ML algorithms (Wang et al., 2021). Early identification of plant diseases is crucial for preventing transmission to other plants, which can result in significant economic losses. The effects of plant diseases can range from minor symptoms to complete plantation loss, making them a major threat to the agricultural economy. Therefore, the key focus of this work is to develop a disease detection system using deep learning techniques.

The rest of this paper is organized as follows: Section 2 elaborates on existing related works done in this domain, research motivation and research contributions. Following this, Section 3 provides a detailed overview of the methodology proposed for this research. Additionally, Section 4 discusses the dataset statistics and presents the experimentation results. Finally, the paper concludes with Section 5.

2. Related works

In this section, various studies on the classification of citrus plant diseases were reviewed. These studies utilize both machine learning and deep learning techniques, which are discussed and analyzed in the subsequent subsections.

2.1. Machine learning methods

Machine learning methods have the capability to effectively identify trends and patterns by analyzing multiple datasets. In recent years, several ml studies have been published on detecting citrus plant diseases. For instance, algorithms such as K-Nearest Neighbor (KNN), K-Means algorithm, and Support Vector Machines (SVM) have been employed for disease classification in citrus plants.

Febrinanto et al. employed the K-Means algorithm for image segmentation to identify diseases on citrus leaves and used the KNN algorithm classify the images into different disease classes (Febrinanto et al., 2019). Safdar et al. proposed a five-step process for the classification of citrus plant diseases, utilizing M-SVM for classification and watershed segmentation for extracting infectious regions. In addition, nine different classification methods were also analyzed, including medium tree, logistic regression, linear SVM, cubic SVM, quadratic SVM, cubic KNN, weighted KNN, ensemble subspace KNN, and ensemble boosted tree (Safdar et al., 2019).

Different methods have been proposed for plant disease detection. For instance, Aidoo et al. employed a maximum entropy model to assess the probability of african citrus triozid occurrence across various habitats worldwide to support the sustainable management of citrus greening disease (Aidoo et al., 2022). Ali et al. presented an automated method for detecting citrus diseases using color histograms, textural descriptors, and the Delta E color difference algorithm. Feature set dimension reduction was achieved by applying Principal Component Analysis (PCA), which was tested using state-of-the-art classifiers (Ali et al., 2017). Senthilkumar et al. proposed a segmentation and

classification model using a backpropagation neural network and weighted lesion segmentation for the identification of citrus plant diseases (Senthilkumar and Kamarasan, 2019). Developing on a previous work, Senthilkumar et al. employed another neural network-based approach for citrus disease classification. Feature extraction was performed using Hough transform and disease classification using the rough fuzzy artificial neural network model (Senthilkumar and Kamarasan, 2020).

Although ML techniques have been applied for citrus disease classification, researchers have shifted towards DL methods to improve results. This is due to ML algorithms being limited by small dataset sizes and requiring more data for accurate classification than DL techniques. Although the issue of limited dataset size also exists in deep learning, numerous data augmentation techniques are available to address this concern. Furthermore, unlike machine learning, deep learning techniques extract features automatically, removing the need for domain expertise in selecting features.

Some of the works have compared the performance of ML and DL models for the classification of citrus diseases. Kaur et al. presented a method for detecting citrus greening disease and evaluated various ML models, finding that the SVM model using a linear kernel achieved the highest accuracy compared to other techniques (Kaur et al., 2020).

2.2. Deep learning methods

Recent research studies have focused on utilizing DL methods for the detection of citrus plant disease. Pre-trained models in combination with transfer learning have been widely applied in this case to classify citrus plant diseases. The use of several deep neural network architectures like EfficientNet, AlexNet, VGGNet, ResNet, YOLO-v4 etc. has gained traction in recent years for citrus plant diseases classification.

For instance, Anwar et al. suggested a comprehensive learning strategy to detect citrus tree disease automatically. Pre-trained weights extracted from various datasets was used to train the DenseNet-121 model, which classified citrus disease images into five distinct categories (Anwar and Anwar, 2022). Tiwari et al. presented a Convolutional Neural Network (CNN) framework based on the DenseNet-101 architecture to perform multiclass plant disease detection while also distinguishing between different types of plant leaves (Tiwari et al., 2021). Recent studies have demonstrated that the You Only Look Once (YOLO) v4 network can achieve faster processing rates and greater accuracy in the classification of citrus plant diseases. Song et al. proposed a method for automatic detection and recognition of citrus diseases using the YOLO-v4 algorithm, employing techniques such as mosaic data augmentation, rotation, and noising to enhance performance (Song et al., 2020). Additionally, Zhang et al. proposed an optimized YOLO-v4 network for detecting citrus diseases and utilized the EfficientNet network for classifying the detected fruits (Zhang et al., 2022). Adeem et al. proposed a DL-based approach to identify and classify citrus diseases effectively using the EfficientNet-B3 model (Adeem et al., 2022). Atila et al. also employed the EfficientNet architecture and conducted a comparative analysis of its performance with other state-of-the-art models to classify plant leaf diseases (Atila et al., 2021). Similarly, Khan et al. introduced a DL-based approach using MobileNet-v2 and SqueezeNet models to classify six distinct citrus diseases. Feature optimization was performed using the WOA algorithm (Khan et al., 2021). Additionally, Sutaji et al. employed an ensemble of MobileNet-v2 and XceptionNet to detect plant diseases through mobile devices (Sutaji and Yildiz, 2022). These studies highlight the diversity of approaches and models that can be used for citrus disease detection through DL-based methods.

Furthermore, various pre-trained models, such as AlexNet, Inception-v3, ResNet-50, ResNet-101, VGG-19, and modified versions of these models, have been evaluated for citrus disease identification and classification. Elaraby et al. utilized pre-trained models such as AlexNet and VGG-19 for discriminative feature extraction from the image

sources and evaluated the performance of SGD with momentum optimizers for citrus disease detection (Elaraby et al., 2022). In another study, Sudharshan et al. conducted a comparative analysis of various neural networks, including ResNet-50, ResNet-101, and VGG-19, for disease classification. The results concluded that VGG-19 offered better accuracy in comparison to the ResNet models (Sudharshan Duth and Bhat, 2022).

Some studies have combined DL with ML methods for automated citrus disease detection. For instance, Senthilkumar et al. proposed an approach that utilized ResNet with an Inception-v2-based feature extractor after an Otsu-based segmentation process and a Random Forest (RF) classifier for disease categorization (Senthilkumar and Kamarasan, 2021). Indra et al. also proposed a DL approach for citrus disease classification that utilized AlexNet for feature extraction and a RF classifier, while Adaptive Gamma Correction (AGC) was applied to enhance the contrast of the input image (Indra et al., 2021). Rehman et al. implemented a DL network that incorporated pre-trained models, MobileNet-v2 and DenseNet-201, for classifying citrus plant diseases. The extracted features from both models were fused and the feature selection process was conducted using WOA (Zia Ur Rehman et al., 2022).

Transfer learning (TL) is a commonly used deep learning method, where pre-trained model weights are transferred to a new classification problem. This approach has been widely utilized in various studies for detecting instances of plant diseases using TL-based techniques. For example, Saini et al. designed a system that incorporated DL techniques and image processing to enable rapid detection in citrus crops. A CNN was employed using TL along with an android application to diagnose the citrus images (Saini et al., 2021a). Similar to previous methods, Saini et al. evaluated various pre-trained models such as Inception-v3, ResNet-50, VGG-16 and VGG-19 for citrus fruit disease identification and classification. The experimental results depict that VGG-19 showed better results and it was implied that TL provided better prediction results with minimal computational resources (Saini et al., 2021b). Yasmeen et al. utilized a modified genetic algorithm to select the best features from input images for citrus disease classification. Modified versions of ResNet-18 and Inception-v3 were trained using TL for the classification task (Yasmeen et al., 2022).

Furthermore, a range of studies employed a combination of modified deep learning models designed to detect citrus diseases. Janarthan et al. devised a patch-based classification network to detect citrus diseases that employed a deep CNN trained with a Siamese neural network classifier (Janarthan et al., 2020). Wu et al. developed a recognition system using DL and field programmable gate array to detect citrus canker disease. The YOLO-v3-MobileNet model was utilized to reduce computational complexity and enable faster detection of the disease, improving hardware feasibility (Wu and Chen, 2021).

Dananjayan et al. compared various CNN detectors for citrus leaf disease detection, including CenterNet, YOLO-v4, Faster-RCNN, DetectoRS, CascadeRCNN, Foveabox and Deformable Detr. Scaled YOLO-v4 P7 was more accurate and fast in predicting diseases, while CenterNet2 with Res2Net 101 DCN-BiFPN was effective in identifying early stages of citrus leaf illnesses (Dananjayan et al., 2022). Furthermore, Generative Adversarial Networks (GANs) have been employed for citrus plant disease classification for better modeling of the data. For instance, Zeng et al. compared the performance of six DL models including VGG, Inception-v3, AlexNet, ResNet, SqueezeNet, and DenseNet, to classify HLB-infected citrus into severity levels. Inception-v3 outperformed the other models, and its accuracy was further improved by using DCGAN-based data augmentation on the best model (Zeng et al., 2020).

Recent studies on citrus plant disease classification have shifted towards using custom CNN architectures, which offer several advantages over multilayer neural networks. For example, CNNs can learn multiple layers of feature representations, handle complex images, and consider local context information, resulting in superior prediction accuracy. Additionally, CNNs can reduce pixel values for efficient use of computing resources without significant information loss, making them

suitable for digital image analysis.

Several recent studies have proposed custom CNN architectures for detecting citrus diseases. Khattak et al. developed a custom CNN architecture with two convolutional and max pooling layers and a flattening layer to extract distinctive features from input images (Khattak et al., 2021). Meanwhile, Çetiner et al. utilized a custom CNN architecture to identify and classify three different citrus diseases and compared it with MobileNet-v2 and ResNet-50 for experimental analysis, achieving better results (Çetiner, 2022). Rahman et al. presented a two-stage CNN with faster R-CNN. The network consisted of four components namely, feature extractor, RPN, ROI pooling, and classifier for the prediction of citrus disease. The model utilized ResNet-101 as the feature extractor, benefiting from residual connections to enable deeper layers and better accuracy without gradient vanishing problems (Syed-Ab-Rahman et al., 2021). Mzoughi et al. proposed the use of semantic segmentation based on DL techniques to identify regions affected by disease in plant leaves and accurately classify them based on the specific disease (Mzoughi and Yahiaoui, 2023). Lastly, Kaya et al. proposed a multi-headed DenseNet model for plant disease identification that combines color and texture features extracted from digital images using various techniques, such as PCA, discrete wavelet transform, and grey-level co-occurrence matrix (Kaya and Gürsoy, 2023). Table 1 summarizes the analysis of few key research works that are reviewed as part of this work.

2.3. Research gaps and motivation

The proposed work addresses the following research gaps for the classification of citrus plant diseases:

1. Many research works use networks that extract either global or local features, resulting in significant feature loss and lower performance rates. This can be addressed by employing feature fusion mechanisms.
2. As DL models become more complex, their subsequent layers contain more parameters which can decrease model efficiency and accuracy. To address this, many models use low-resolution feature maps that require up-sampling data from one layer to the next.
3. Existing works often employ high-level features which are generally extracted from the later layers of the network, leading to poor model generalization during training. Hence, feature representations from different levels of abstraction need to be processed for precise classification.
4. Few works have adopted either spatial or channel attention mechanisms in citrus disease classification tasks to improve model performance. An efficient attention module that seamlessly combines both spatial and channel attention mechanisms is essential to improve the accuracy and precision of citrus disease classification.

2.4. Research contributions and novelty

The key contributions of the proposed work are as follows:

1. The proposed network integrates features from two distinct tracks that extract both global and local features. This helps the network to better contextualize the learned features and improve the precision of classification without losing important information.
2. The proposed network contains a Swin transformer track that utilizes hierarchical feature maps to enable multi-scale feature learning. This approach allows the network to maintain a lesser number of parameters, which in turn contributes to better model performance.
3. A Feature Pyramid Network (FPN) block was integrated at the end of the CNN track to fuse high-level and low-level features extracted by the CNN blocks to improve the generalization of the model.
4. The Shuffle Attention (SA) module in the proposed network employs both spatial and channel attention mechanisms to capture

Table 1

A summary highlighting the strengths and weaknesses of existing works on citrus disease detection.

S. No.	Source	Method	Strengths	Limitations	Accuracy (in %)
1	Febrinanto et al., 2019	Implemented a model for citrus disease classification that used KNN algorithm to classify the images into different disease classes.	Image segmentation using K-Means was performed to remove noise.	Low class-wise accuracy for the healthy leaf class (63.33%) which may potentially affect overall accuracy, especially when majority of leaves in a sample are healthy.	90.83%
2	Safdar et al., 2019	Presented a five step process for citrus disease classification. SVD based feature extraction and PCA reduction were performed while M-SVM was utilized for classification. 9 different classification models were analyzed and compared.	Performed pre-processing techniques like noise reduction and contrast enhancement to improve image quality. Implemented watershed segmentation to help extract infectious regions.	Watershed-based segmentation has limitations due to parameter selection, which can vary depending on image characteristics, making it challenging to find a universal solution.	Combining 3 datasets (Citrus Image Gallery, Plant Village, and self-collected) Classification accuracy - 95.5%
3	Senthilkumar and Kamarasan, 2019	Proposed a back propagation neural network for citrus identification.	Performed contrast enhancement of images and weighted lesion segmentation for improved model performance.	Insufficiency in the number of training samples or the number of weak classifiers may negatively impact the accuracy and generalizability of the final classification model. May not be suitable for real-time applications or resource-constrained devices due to its computationally expensive processes such as bilateral filtering, optimal weighted segmentation, hough transform.	96.20%
4	Senthilkumar and Kamarasan, 2020	Implemented a rough fuzzy artificial approach for citrus disease classification. Hough transform was utilized to perform feature extraction.	Bilateral filtering was performed to remove the background noises and improve image quality.	Small size of the dataset used (150 images) without any reported use of augmentation techniques, may limit the model's ability to generalize to new and diverse data.	96.93%
5	Kaur et al., 2020	Performance of various ML models was evaluated. SVM classifier with linear kernel variant outperformed other techniques.	Used a combination of high boost and morphological filters with varying structuring elements to improve image quality by reducing noise and enhancing contrast. Genetic algorithm based feature optimization, lead to a significant accuracy improvement	Traditional ML often involves manual feature extraction, where experts manually design algorithms to extract relevant features from the input data.	SVM classification with linear kernel variant – 90.4% (best accuracy) Naïve Bayes classification with Gaussian kernel – 89.9% Decision tree classification with fine kernel – 85.7% KNN classification with medium kernel variant – 86.3%
6	Anwar and Anwar, 2022	Implemented DenseNet-121 model on pre-trained weights via transfer learning for citrus disease detection.	Utilized pre-trained weights via extracted from various dataset via transfer learning. Reducing the need for extensive training on large datasets.	Different source and target domains may cause hindrance to the model performance when transfer learning is employed.	92%
7	Tiwari et al., 2021	Deep framework based on DenseNet-101 architecture was utilized to classify different types of plant leaves and identify plant diseases.	Model can be used to operate in real-time and can be extended to integrate with camera-based systems.	The small size of the dataset used may lead to overfitting and reduced generalizability of the model.	85.55%
8	Zhang et al., 2022	Proposed a YOLO-v4 network for disease detection and utilized EfficientNet-B4 network for optimum fine classification.	Utilized two separate models for disease detection and classification.	The detection network and fine classification network may be integrated into a unified framework to streamline the training procedures.	89% and 87.2% respectively
9	Elaraby et al., 2022	Employed pre-trained models such as AlexNet and VGG-19 for discriminative feature extraction using transfer learning.	Utilized pre-trained models via transfer learning, reducing the need for extensive training on large datasets.	While transfer learning can be an effective technique, it heavily relies on the similarity between the source and target domains. Simplifying the neural network to only one layer for the classifier may limit its ability to capture complex patterns and features, reducing classification accuracy.	94.30%
10	Janarthan et al., 2020	Employed a patch-based classification for citrus disease using a deep CNN.	Siamese neural network classifier was utilized for effective classification on small dataset.	CCL'20 dataset consists of optical RGB citrus leaf images with prominent symptoms, which may not be adequate	95.04%
11	Dananjayan et al., 2022	Implemented and fine-tuned CNN-based detectors such as CenterNet, YOLO-v4, Faster-RCNN, DetectoRS,	Proposed a new annotated dataset (CCL'20). Scaled YOLO-v4 resulted in fast disease	Avg Recall - 92.8% Avg Precision - 98%	

(continued on next page)

Table 1 (continued)

S. No.	Source	Method	Strengths	Limitations	Accuracy (in %)
12	Zeng et al., 2020	Cascade-RCNN, Foveabox, and Deformable Detr for detecting citrus leaf diseases.	prediction, while using CenterNet2 with Res2Net 101 DCN-BiFPN was effective for early detection of citrus leaf diseases.	for early detection of newly infected citrus without visible symptoms, as reported in the paper.	Inception V3 (epochs = 60): Accuracy - 74.38% Inception V3 with DCGANs: Accuracy - 92.60%
13	Khattak et al., 2021	Various pre-trained models were compared against each other for citrus disease classification and Inception-v3 outperformed the other models.	DCGAN based data augmentation was performed which helped boost the model accuracy by 20%.	Different source and target domains may cause hindrance to the model performance when transfer learning is employed.	94.55%
14	Çetiner, 2022	Another custom CNN was employed to classify citrus fruits and leaves diseases.	Model exploited 2 pairs of convolution and max pooling layers and a flattening layer to extract complementary discriminative features.	The limited number of images in the dataset (only 213) may increase the risk of overfitting and hinder the ability of the model to generalize well to new and unseen data.	96%
15	Syed-Ab-Rahman et al., 2021	Proposed a custom CNN architecture for citrus disease classification. In addition, the model was compared with MobileNet-v2 and ResNet-50 for experimental analysis and achieved better results.	Detected ROI region and borders on the image. HSV color space was used during image pre-processing to reduce the loss of information.	The small size of the dataset used may lead to overfitting and reduced generalizability of the model.	94.37%
		Employed a novel 2-stage classification model based on Faster R-CNN and utilized ResNet-101 as the feature extractor.	Used residual connections to enable deeper layers and better accuracy without gradient vanishing problems	Basic augmentation has been reported to balance the class size. The small size of the dataset used may lead to overfitting and reduced generalizability of the model.	

informative features from different spatial locations and channels of the input feature maps. This enables the network to selectively emphasize the most relevant features, thereby enhancing the discriminative power of the network.

3. Proposed network

The proposed network combines features extracted from two distinct tracks: the GSDFP track and the Swin transformer track. The GSDFP track is a CNN-based method that extracts local features from input images, while the Swin transformer track captures global features. By fusing these two sets of features, the network can train a more robust model resulting in higher classification accuracy. A visual representation of the overall process flow of this approach is depicted in Fig. 1. The

methodology followed several main steps which include data pre-processing, data splitting, data augmentation followed by training of the deep neural network and then performance evaluation and classification.

The main motivation behind using the GSDFP track is to demonstrate a network design that takes advantage of three main components: 1) Group Shuffle Depth Wise (GSDW) blocks, 2) 2D conv and Batch Normalization (BN) layers, and 3) FPN block. The Swin transformer track consists of multiple stages of the Swin Transformer block in combination with linear embedding and patch partition in a linear fashion. The idea behind the Transformer track is to leverage the hierarchical design and shifted window technique of the Swin Transformer to perform better feature accumulation, which can boost the overall accuracy and learning parameters in the network. The features from the

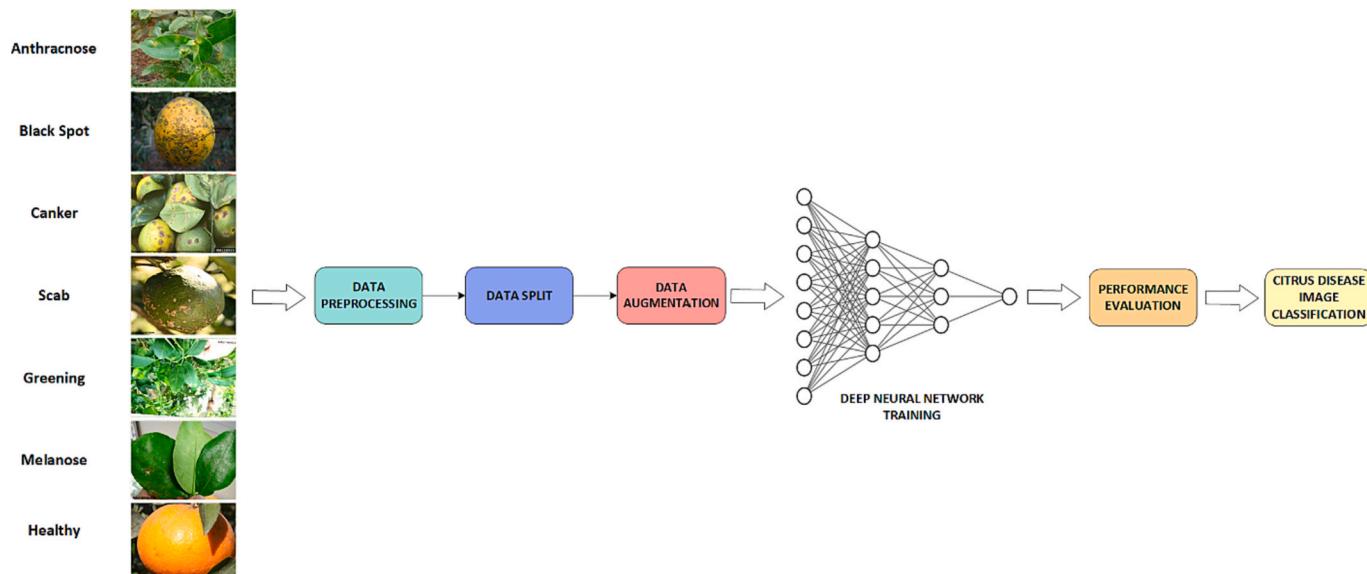


Fig. 1. Schematic representation of the Proposed Approach.

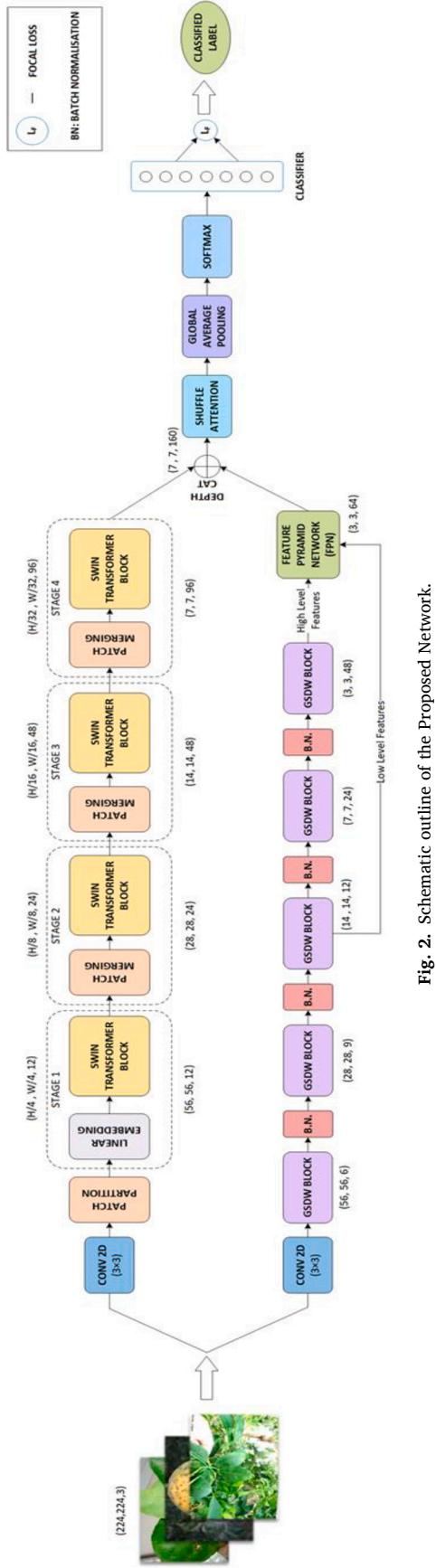


Fig. 2. Schematic outline of the Proposed Network.

two branches are fused using depthwise concatenation. Subsequently, the network utilizes various feature weighting and reduction modules, such as SA and GAP. The final classification layer employs softmax to classify the input data, with focal loss serving as its loss function.

A detailed design of the proposed network is presented in Fig. 2. The steps involved in this schematic are as follows: First the input images are preprocessed and resized to a fixed dimension of $224 \times 224 \times 3$. These images are then fed into two parallel tracks, namely the GSDFP track and the Swin transformer track. Both tracks consist of various modules that extract multiple parameters and refine the network's ability to detect citrus plant diseases in the images. At the end of each track, the extracted parameters are fused together using the depthCat network. To introduce variability to the features, an SA module is employed. Following this, the most significant parameters are pooled together using the GAP reduction module. Finally, the output from the last layer of the network is combined with a focal loss function to balance the class distribution and produce the final output.

3.1. Swin transformer track

This section describes the working of the Shifted WINdows (SWIN) transformer module to improve multi-scale feature learning. The Swin transformer module achieves this by calculating self-attention only for local windows. The module is structured to generate hierarchical feature representation through four stages that occur sequentially. The architectural outline of the Swin transformer track is illustrated in Fig. 3. This track involves incorporating preprocessing modules for preparing the image prior to inputting it into the transformer. The transformer consists of several nested modules, each with varying input sizes, resulting in the generation of a feature map as output.

The Swin transformer block processes input RGB images through a 2D convolution block and extracts relevant features. It is followed by the patch partition, where the image is split into non-overlapping patches as tokens. The feature of each patch is represented as a concatenation of the RGB values of individual pixels, and the patch size used in the implementation is $H \times W$. The feature dimension of each patch is $H \times W \times 3$, and a linear embedding layer is applied to reduce the dimensionality of the patch embeddings and project them into an arbitrary dimension. The dimensions of the patches are determined by the patch size hyper-parameter, denoted as P . The patch size represents the spatial extent of each patch in the input image. With a value of P set to 4, each patch becomes a square region of size 4×4 pixels. The patch embeddings are then processed by the Swin transformer block through a series of sub-units, each consisting of a normalization layer, an attention module, another normalization layer, and an MLP layer. Fig. 4 provides a detailed representation of the interaction between two successive Swin transformer blocks. The initial sub-unit uses a Window Multi-Head Self Attention (W-MSA) module to capture the dependencies between the patch embeddings. The next sub-unit uses a Shifted Window Multi-Head Self Attention (SW-MSA) module to attend to shifted windows of patch embeddings, which helps the model capture long-range dependencies with fewer parameters (Swin Transformer, 2022). The output of the second sub-unit is then passed through a skip connection to the input of the Swin Transformer block, and the hierarchical feature representation is employed to capture the long-range dependencies. Between each block of layer normalization with multi head self-attention and layer normalization with MLP, element-wise addition of parameters is performed. The down-sampling module reduces the spatial resolution of the feature map while increasing the number of channels. The patch merging layer is applied at the end of each stage to concatenate the features of each set of $n \times n$ neighboring patches depth-wise to down-sample feature maps by a factor of n . Finally, the output layer of the Swin Transformer track has 96 channels, and the feature map obtained from this track is combined with that of the CNN track to generate a complete feature map that is used for training the classification model.

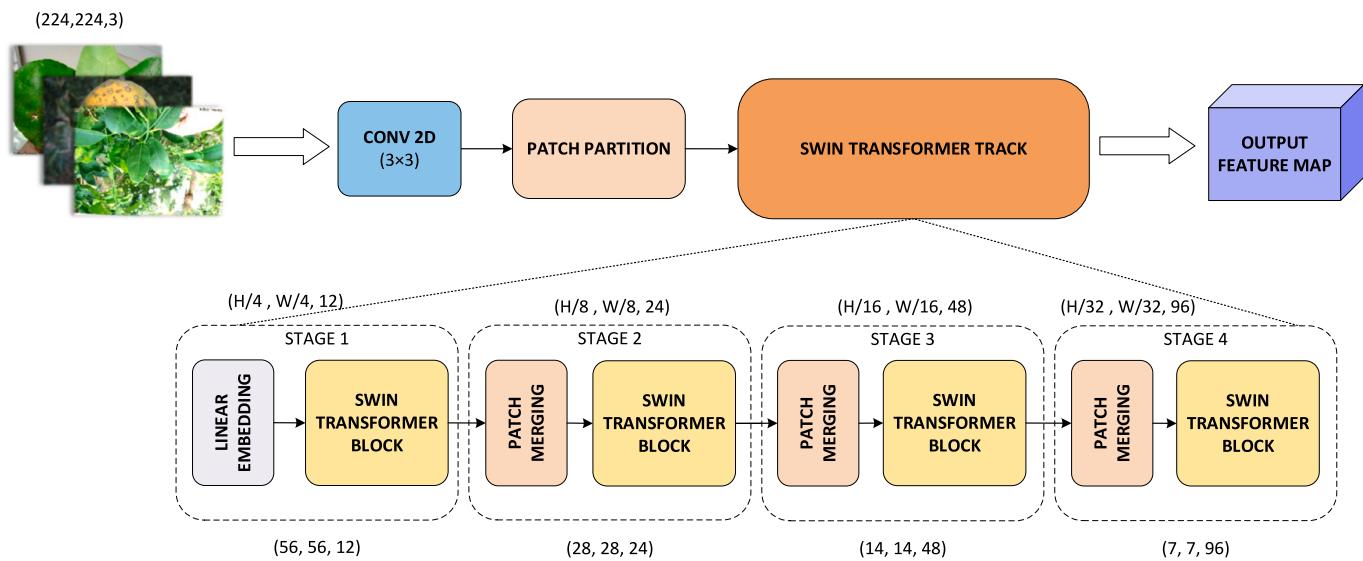


Fig. 3. Architectural overview of the Swin transformer track.

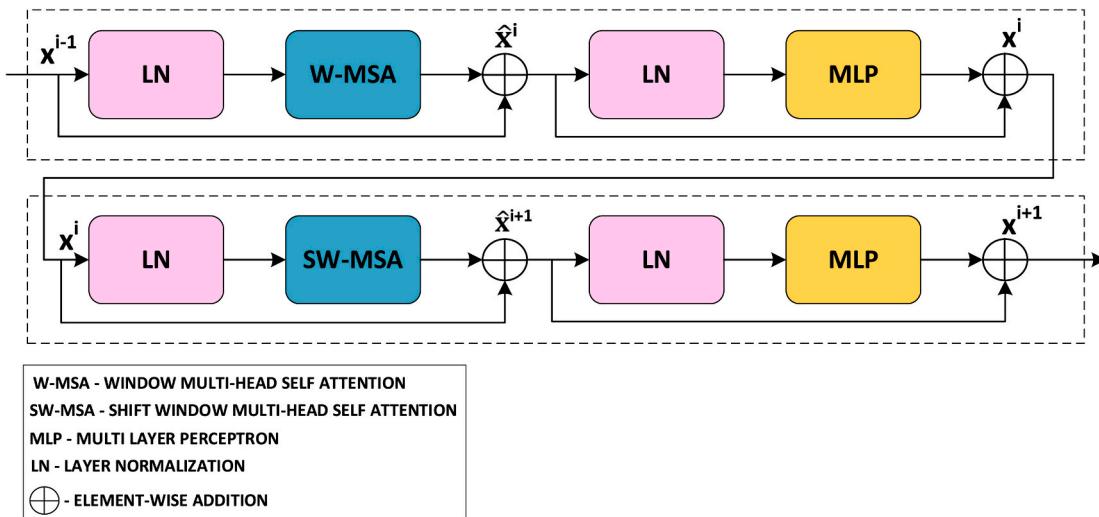


Fig. 4. Interaction between two successive Swin transformer blocks.

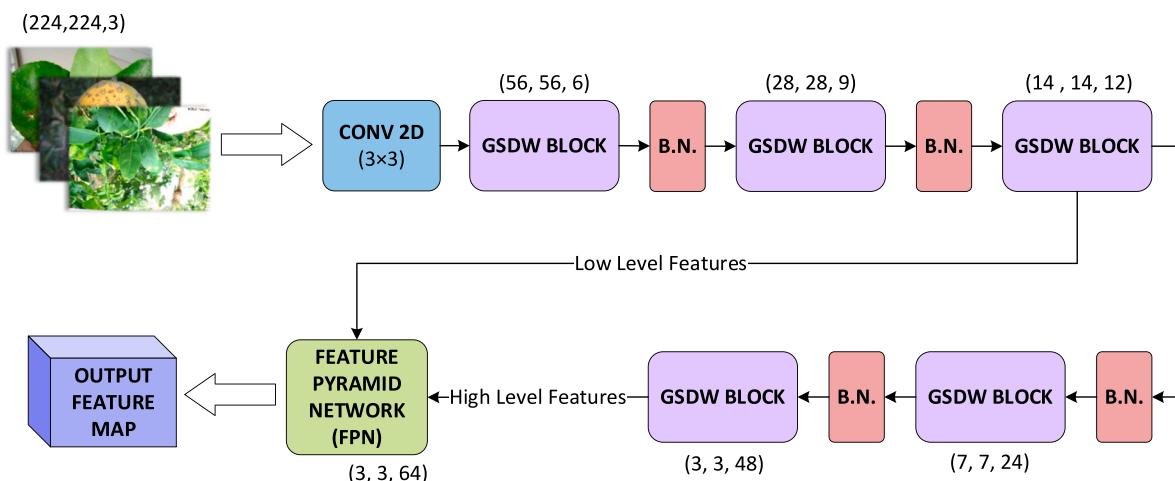


Fig. 5. Schematic overview of the GSDFP track.

3.2. Group shuffle Depthwise feature pyramid track

This section highlights the role of the GSDFP track in the proposed network, as depicted in Fig. 5. The GSDFP track works in parallel with the Swin transformer track for dual multi-level feature extraction. The network operates on input dimensions of $224 \times 224 \times 3$. The feature depth of the GSDFP track increases twofold through the use of CNN, resulting in 64 channels at the penultimate layer. The extracted features from both tracks are then combined and subjected to further processing to train the model for accurate classification of citrus plant diseases.

The GSDFP track is composed of a series of alternating modules that include 2D convolutional layers, BN and GSDW blocks. The final block of this track is the FPN block, which is utilized to fuse the extracted features to enhance the model's robustness and generalization ability. Low level features from the center of the track are added to the output feature map along with high level features from the final GSDW block with the help of the FPN block, which also combines all the network parameters.

Fig. 6. illustrates the GSDW module, which enhances local feature extraction by exploiting Group Convolution (Gconv), channel shuffle, and Depth-Wise Separable Convolution (DWSC) (Li and Li, 2022). These modules are placed in a sequential manner. The Gconv divides the input tensor into multiple groups and performs convolution separately on each group, reducing the computational complexity of convolutional layers. The module also uses channel shuffle to rearrange the channel dimensions of the input tensor, allowing cross-group information flow and enhancing the network's capacity to extract global features. Finally, the proposed network incorporates the DWSC module, which has two important roles. First, it reduces the number of nonlinear operations in the network, while maintaining expressive power. Second, it significantly reduces the computational cost of the network. DWSC accomplishes this by decomposing the convolution operation into two separate steps: depth-wise convolution and point-wise convolution. This reduction in the number of parameters makes the network more efficient and suitable for resource-constrained scenarios or when dealing with large-scale datasets.

Following each GSDW block, BN was applied to normalize the layer-wise inputs through re-centering and rescaling, thereby improving the stability and speed of network training, as described in Eq. (1).

$$Z^N = \left(\frac{z - m_z}{s_z} \right) \quad (1)$$

where, m_z represents the mean of the neurons' output and s_z represents the standard deviation of the neurons' output.

The FPN block serves as the final component of the GSDFP track, which leverages lateral connections to integrate semantically rich features from low-resolution and high-resolution feature maps. After extracting features from both the GSDFP and Swin transformer tracks, the network performs depthwise concatenation operation to merge the features from both tracks. By learning from two distinct feature extraction tracks, the network can effectively leverage their complementary information to improve better feature representation and enhance the accuracy of end predictions.

Using FPN in image classification tasks, the model can effectively capture both low-level and high-level features in the images. This can improve the accuracy of the classification task (Lin et al., 2017). Furthermore, FPNs in this case help to enhance model robustness to image variations by leveraging multi-scale feature maps (Ghiasi et al., 2018). Therefore, it can be considered a valuable addition to the proposed work, even though it is primarily used in object detection tasks.

3.3. Shuffle attention block

The Shuffle Attention (SA) module is designed to enhance feature interaction between the primary and spanning branches of the network. It is employed after depthwise concatenation, which combines features from the two tracks to create a comprehensive feature map. The SA module has been shown to improve the model's training and performance significantly. The schematic outline of the SA block is illustrated in Fig. 7, and it involves operations such as grouping, splitting, and fusion of features. The module works by breaking the input feature map into G groups, where G is the number of groups specified. Each group contains a subset of channels, calculated by dividing the total number of channels C by G , resulting in feature maps with C/G channels per group. After feature grouping, the module applies both channel attention and spatial attention layers to each group independently. The channel attention layer (X^C) learns the importance of each channel within a group, generating a channel attention map of size $C/G \times 1 \times 1$. Similarly, the spatial attention layer (X^S) learns the importance of each spatial location within the group, producing a spatial attention map of size $H \times W \times 1$, where H and W represent the spatial height and width, respectively. The next step involves aggregating the channel attention map (X^C) and the spatial attention map (X^S) for each group. This is accomplished by element-wise multiplication, resulting in an intermediate feature map X' with size $R^{C \times H \times W}$, where each channel's attention has been combined with its corresponding spatial attention. To further enable information transfer between different sub-features within the group, a "channel shuffle" operator is employed. This operator allows the module to shuffle and concentrate attention on minute details in the image. Throughout the process, the SA block uses sigmoid and weight balancing functions at various stages, along with other operations like concatenation and element-wise multiplication, to refine and control the attention mechanisms. Overall, the Shuffle Attention module enhances feature interaction through the use of channel and spatial attention layers applied to grouped feature maps. By effectively learning the dependencies between channels and focusing on important spatial locations, the module produces an enhanced feature map that improves the overall performance of the model in processing complex visual information (Zhang and Yang, 2021).

3.4. Classification

The feature vectors obtained from the SA block are subjected to global average pooling to decrease the model parameters in the network, as illustrated in Fig. 2. The final layer of the network employs a softmax

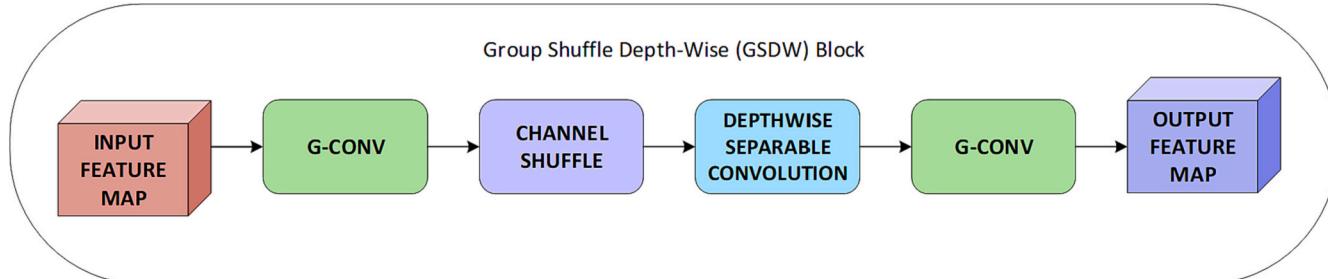


Fig. 6. Schematic overview of the GSDW block.

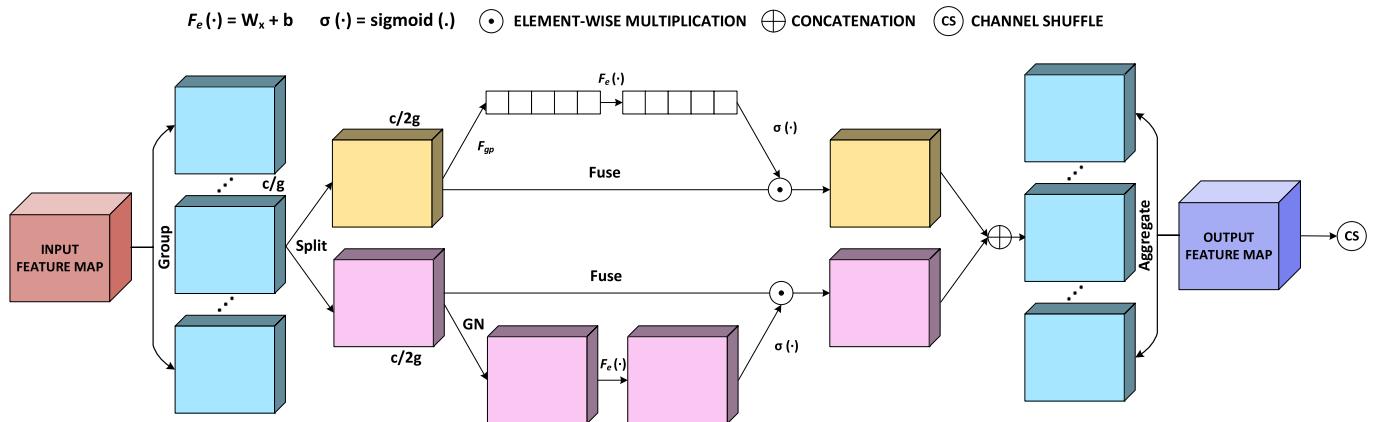


Fig. 7. Schematic overview of the SA block.

function for multi-class classification of the citrus plant disease images. To address the class imbalance in the data, the proposed approach employs the focal loss function, which has been found to be an effective method for improving the model's robustness by training it on classes with fewer images. This function addresses the class imbalance issue by assigning lower value to easy samples and emphasizing challenging ones. The network parameters are learned by maximizing the focal loss

of the predicted class probabilities with respect to the target class, as shown in Eqs. (2) and (3).

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} \log p_t \quad (2)$$

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (3)$$

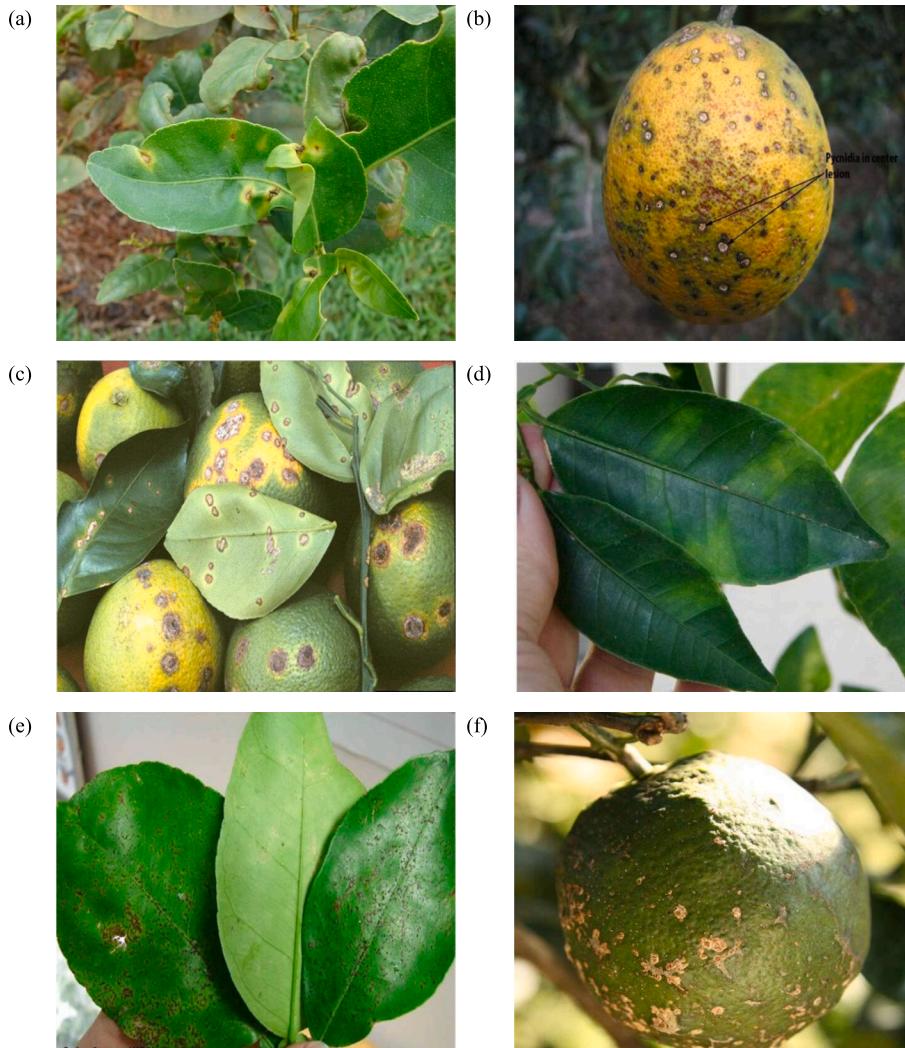


Fig. 8. Sample images from the combined citrus dataset (a) Anthracnose (b) CBS (c) CC (d) Greening (e) Melanose (f) Scab.

where, $(1 - p_t)^\gamma$ is the modulating factor, γ is focusing parameter, α_t are the class weights, y is the true class label and p is the class probability whereas p_t is the model's estimated probability (Lin et al., 2020). As a result, the focal loss is a cross-entropy loss that is dynamically scaled, with the scaling factor decrementing to zero as the confidence in the correct class increases. This function is more effective than the standard cross-entropy loss, particularly in cases with a significant class imbalance.

4. Results and discussion

This section presents a summary of the dataset used in the research, along with the data augmentation techniques employed. Additionally, it describes the environmental setup and extensive ablation experiments performed to assess the influence of individual model components on the overall performance. Finally, the performance of the proposed approach is analyzed comprehensively using quantitative and qualitative evaluations to demonstrate its effectiveness.

4.1. Dataset description

This research work utilizes a dataset composed of images of healthy and diseased citrus fruits and leaves, combined from two sources: the Citrus Disease Image Gallery Dataset (Image Gallery, 2023) and the Citrus Plant Dataset (Rauf, 2019). The dataset comprises seven main categories, namely Anthracnose, Citrus Canker (CC), Citrus Black Spot (CBS), Citrus Greening, Melanose, Sweet Orange Scab, and Healthy, as depicted in Fig. 8.

4.2. Data augmentation

We have discussed the data augmentation techniques employed to prepare the images for training in this subsection. The combined citrus image dataset consists of 842 images with varying dimensions. To facilitate efficient feature extraction, each image was resized to a standardized resolution of 224×224 pixels. The dataset was then split into three parts - 60% allocated for training, 20% for validation, and the remaining 20% designated for testing. Given the limited size of the dataset, data augmentation was employed to improve the generalizability of the model and reduce the risk of overfitting. The following augmentations were carried out: (1) Random rotation of the images with a rotation range of 45° . (2) Random flipping of the images horizontally. (3) Random flipping of the images vertically. (4) Shifting the image with a width and height shift range of 0.2. (5) Applying a shear transformation with a shear range of 0.7. (6) Adjusting the brightness of the images with a brightness range of [0.5, 1.5]. Table 2. summarizes the class-wise dataset details used for training, validation and testing before and after augmentation of the data.

4.3. Environment setup

All the experiments with the proposed network were executed on a

24GB Nvidia A10G Tensor Core GPU using Pytorch on an AWS EC2 instance. The computing environment for this training consisted of Ubuntu 20.04 operating system, 4 AMD vCPUs, and 16GB RAM. During the training of this model, hyper parameter tuning was performed using the Adam gradient descent optimization algorithm. To minimize the effect of class imbalance in our training dataset, focal loss was chosen as the loss function.

4.4. Hyperparameter tuning

Hyperparameter optimization and analysis plays a critical role in the development of deep learning models as they significantly influence a model's ability to learn and generalize input data. This study aims to enhance the model performance by optimizing three key hyperparameters: (1) dropout rates in all dropout layers, (2) learning rate and weight decay, and (3) gradient update optimization algorithm. The research explores a wide array of feasible values for each hyperparameter, including dropout probability values between 0.4 and 0.7 in increments of 0.05, learning rate and weight decay parameters ranging from 0.0001 to 0.1 in multiples of 10, and two gradient optimizers, namely Adam and Stochastic Gradient Descent (SGD). The outcomes reveal that a dropout rate ranging from 0.5 to 0.6 produces the most optimal model performance. Dropout serves as a regularization technique to mitigate overfitting in neural networks. It achieves this by randomly deactivating a fraction of neurons during training, introducing noise that fosters the learning of more robust and generalizable representations. A higher dropout rate deactivates a larger number of neurons during training, thereby enhancing the regularization effect. This regularization prevents the model from closely memorizing the training data, leading to improved generalization on unseen examples. Moreover, dropout helps counter co-adaptation among neurons, where certain neurons overly rely on specific others. Co-adaptation often leads to overfitting on training data and poor generalization to new examples. By dropping out neurons, dropout disrupts co-adaptation, encouraging neuron independence and robustness to input variations. A larger dropout rate further reduces co-adaptation, promoting better generalization performance. Moreover, the study reveals that the network reaches maximum convergence when the learning rate parameter value of 0.0001. Finally, the results demonstrate that the Adam optimizer consistently outperforms SGD in achieving better outcomes.

4.5. Ablation studies

This section presents the ablation analysis performed to assess the impact of individual components of the proposed model on its overall performance. The primary objective was to determine the optimal combination of model components that would yield the best performance. The findings of this analysis provide valuable insights into the behavior of the model. The proposed network is analyzed using different performance measures like Accuracy, Precision, Recall, and F1-score. Additionally, accuracy and loss curves were presented to interpret the trend of convergence for each network. Along with these plots, the ROC curve and confusion matrix for the proposed system were also analyzed.

Table 2

Summary of the class-wise dataset statistics.

S.No.	Class name	Before augmentation			After augmentation		
		Train set	Validation set	Test set	Train set	Validation set	Test set
1	Anthracnose	6	2	2	73	18	2
2	Citrus Black Spot	127	41	41	1359	340	41
3	Citrus Canker	152	50	50	1645	411	50
4	Citrus Greening	143	48	47	1534	383	47
5	Healthy	48	16	16	561	140	16
6	Melanose	13	4	4	154	39	4
7	Sweet Orange Scab	20	6	6	237	59	6
	Total	509	167	166	5563	1390	166

The results demonstrate a consistent improvement in overall performance upon the incorporation of the proposed modules to the network.

4.5.1. Analysis of the Swin transformer track

In this section, we examine the efficiency of the Swin transformer feature extraction track. The input images are partitioned into tokens, which are then processed through a series of Swin transformer blocks. These blocks use self-attention mechanisms and layer normalization to extract features from the input patches and capture both local and global context information. The four sequential stages of Swin transformer blocks further enhance the hierarchical representation of the input image. The output of the Swin transformer branch is subjected to adaptive average pooling, followed by flattening and linear transformation, before being passed to the classifier. The network was trained using the Adam optimizer for 100 epochs. The observations recorded during the training phase are illustrated in Fig. 9, with an accuracy of 82.28% achieved on the testing dataset. Additionally, the precision, recall, and F1 score metrics were computed, resulting in values of 82.16%, 82.28%, and 82.06%, respectively. In summary, the Swin Transformer network uses token-based processing, self-attention, layer normalization, hierarchical representation, and adaptive average pooling to effectively extract features from input citrus leaf images and achieve high accuracy for the classification task.

4.5.2. Analysis of the GSDFP track

In this section, we analyze the effectiveness of CNN-based feature extraction branch. The input images undergo a sequence of five GDSW blocks before being fed into the feature pyramid network. BN is utilized to facilitate optimal feature learning. The FPN block performs a fusion of features from the second and fifth GDSW blocks. Grouped convolution and depth-wise convolution are employed to reduce parameter size and propagate features across groups. The feature maps in the FPN branch pass through an encoding track, bottleneck, and decoding track. The pooling layer is the final layer, aligning the output of the GSDFP branch with that of the transformer branch. This branch is trained using the Adam optimizer for 100 epochs. The observations recorded during the training phase are illustrated in Fig. 10. The GSDFP track achieves an accuracy of 86.82% on the testing dataset, demonstrating its effectiveness for citrus leaf image classification. This is due to the presence of GDSW blocks, grouped convolution, and depth-wise convolution in the GSDFP track to capture low-level and high-level features from input images. The FPN block fuses features from multiple stages and the encoding, bottleneck, and decoding tracks enhance the feature representation.

4.5.3. Analysis of the proposed system without shuffle attention

In this section, we evaluate the performance of the proposed system without the inclusion of the self-attention (SA) block. The model was trained using the Adam optimizer for 100 epochs, and its effectiveness was compared to the overall architecture with the SA block. The observations recorded during the training phase are illustrated in Fig. 11. The results obtained from training the proposed network without SA are as follows: accuracy of 89.92%, precision of 90.35%, recall of 89.89%, and F1-score of 90.12%. While the proposed network without SA showed notable performance, its absence led to a decline in the overall system's performance. The SA block's unique ability to preserve spatial relationships and capture meaningful features is evidently a significant contributor to the overall effectiveness of the model.

4.5.4. Analysis of the proposed system without focal loss

The aim of this experiment is to analyze the proposed system without using focal loss. As part of this experiment, we introduce a network that combines the feature maps of the GSDFP and transformer branches. Cross-entropy was used as the loss function to fine tune the weights of the network. The observations recorded during the training phase are illustrated in Fig. 12, and an accuracy of 91.69% was achieved for the test set. This is due to the fact that the combination of the feature maps from the GSDFP and transformer branches allows the model to capture more meaningful and informative features, which can contribute to achieving high accuracy.

4.5.5. Analysis of the proposed system with focal loss

In this section, we investigate the effectiveness of using Focal Loss as the loss function for training the overall architecture. The Focal Loss function was applied to tackle the class imbalance issue in the dataset and reduce the impact of easy examples and outliers. The overall architecture was trained using the Adam optimizer for 100 epochs with the Focal Loss function. The observations recorded during the training phase are illustrated in Fig. 13. With the replacement of the loss function, the model resulted in an accuracy of 98.19%. This is because Focal loss is a modified form of cross-entropy loss that assigns higher weights to challenging-to-classify examples. This aids in enhancing the performance of model on the imbalanced test dataset and achieving a high accuracy. Fig. 14 and 15. present the confusion matrix for classification and ROC curve obtained for the proposed network respectively. Based on the ROC curve, it can be interpreted that the performance is nearly optimal, as the curve is close to the top left corner of the plot. This indicates that the classifier has a high true positive rate and a low false positive rate, making it highly accurate in identifying the positive class.

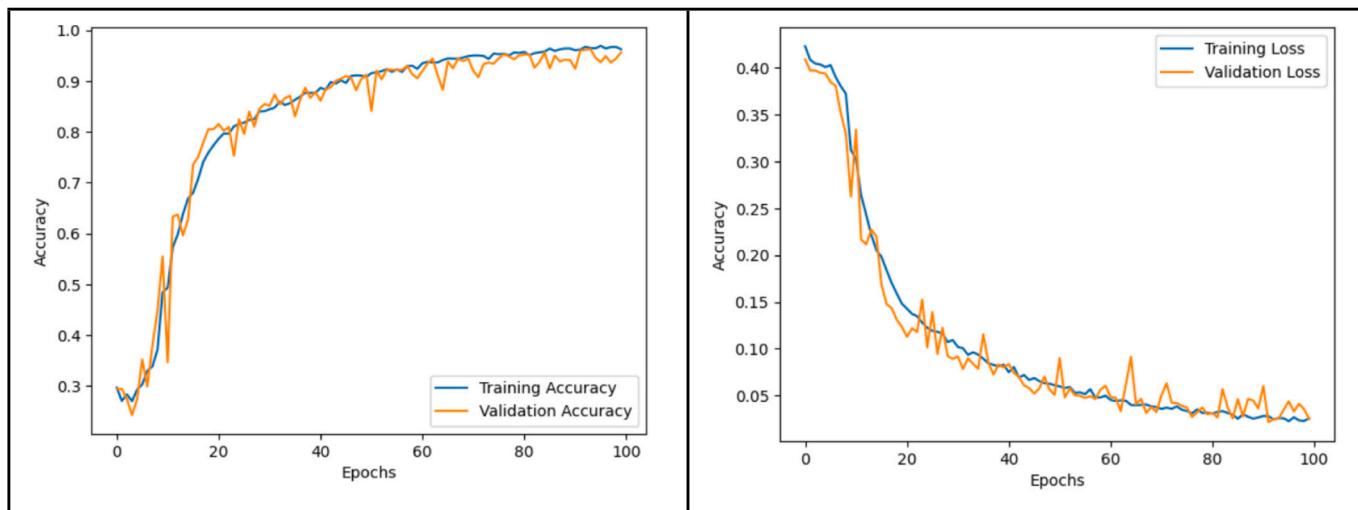


Fig. 9. Analysis of the Swin Transformer track (a) Accuracy (b) Loss.

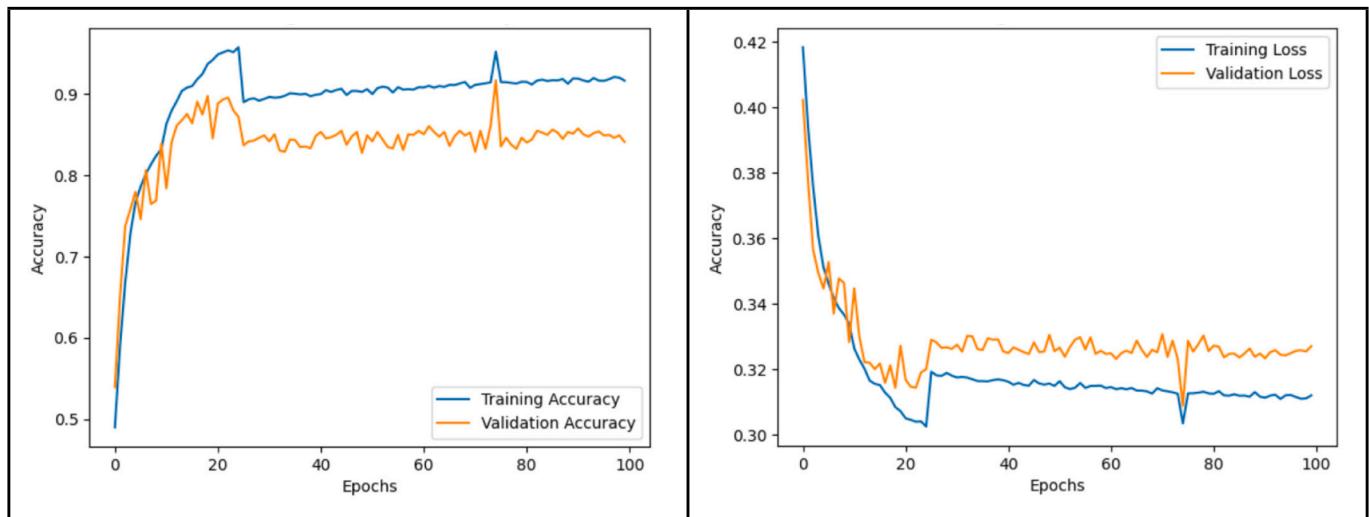


Fig. 10. Analysis of the GSDFP track (a) Accuracy (b) Loss.

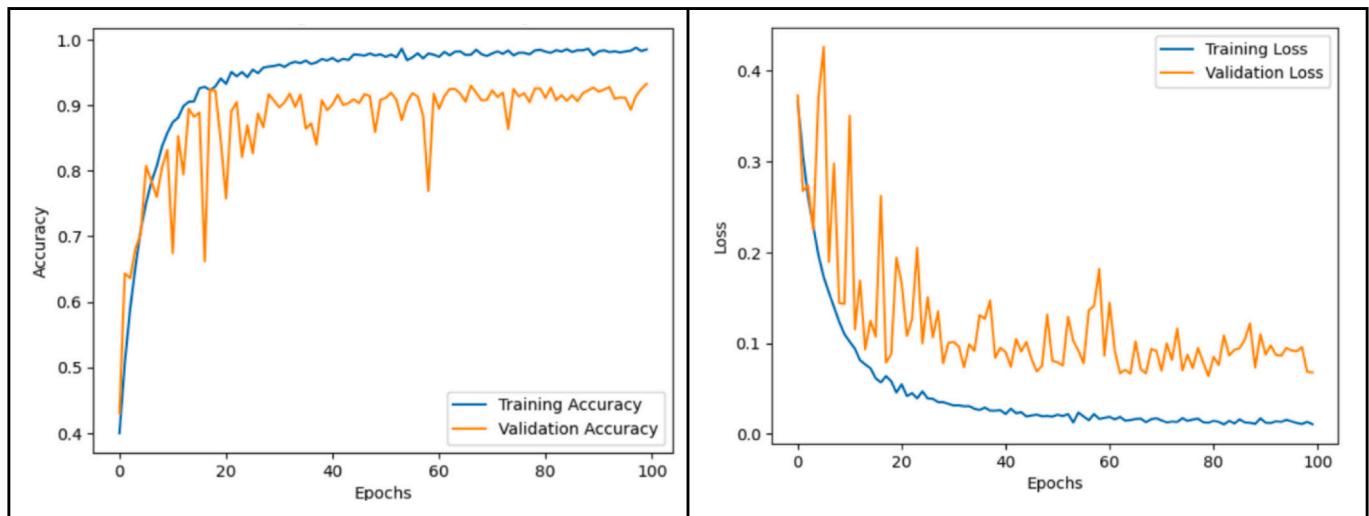


Fig. 11. Analysis of the proposed network without SA block (a) Accuracy (b) Loss.

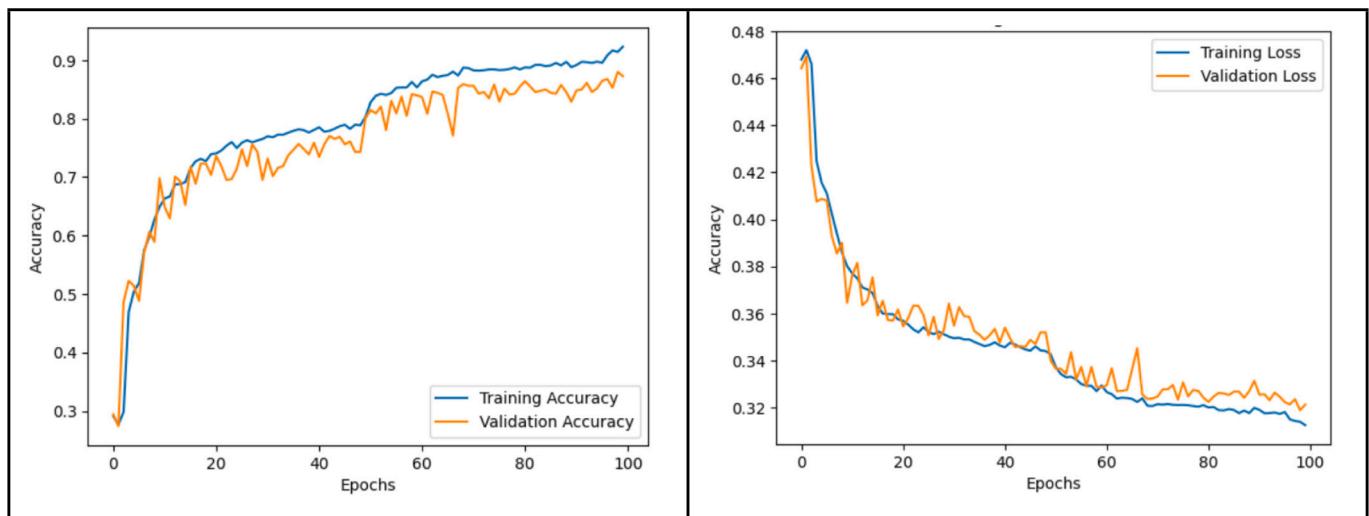


Fig. 12. Analysis of the proposed network without focal loss (a) Accuracy (b) Loss.

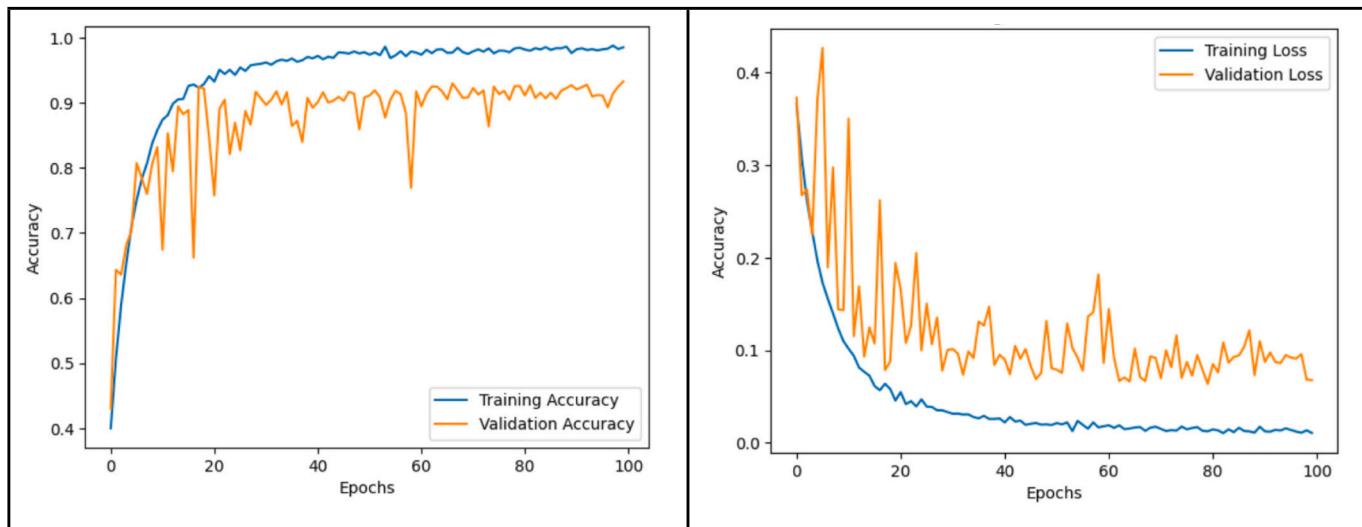


Fig. 13. Analysis of the proposed network (a) Accuracy (b) Loss.

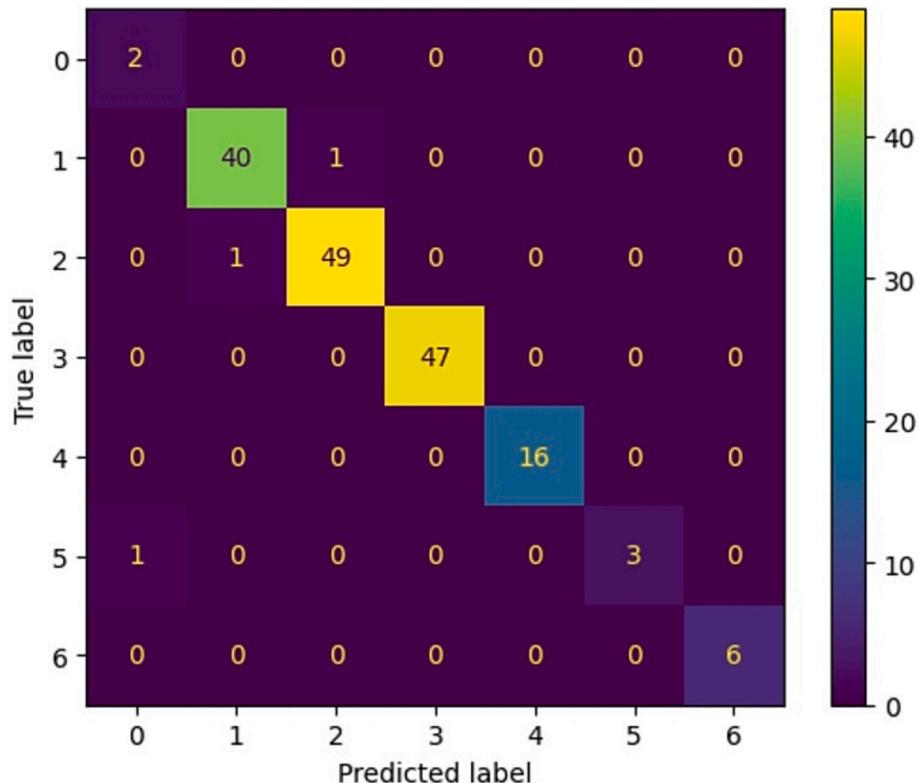


Fig. 14. Confusion matrix for the proposed network.

Also, Table 3 summarizes the ablation studies conducted and its observed metrics. (See Fig. 15.)

4.6. Performance analysis

In this research, a novel DL network was developed for the classification of citrus disease, and its performance was analyzed and compared against existing works reported in the literature. The accuracy, precision, and recall from similar studies are presented in Tables 4 and 5. Tables 4 and 5 compare works that utilize Citrus Disease Image Gallery and the Citrus Plant datasets respectively, ensuring a similar level of complexity across all works. The results showed that pre-trained

architectures achieved accuracies ranging from 89% to 94%. The proposed fusion of local and global contexts significantly improved the accuracy to 98.19%, surpassing the results of earlier works.

To further evaluate the performance of the proposed network, a state-of-the-art comparison was done with commonly used DL models for image classification. Five standard CNN's which include VGG-16, AlexNet, MobileNet-V2, EfficientNet-B0, and DenseNet-201 were reimplemented and trained on the same benchmark dataset as the proposed network. The performance metrics of these models are presented in Table 6, which include accuracy, precision, recall, F1-score, and AUC. The results showed that the proposed network demonstrated greater performance (98.19%) than the other five state-of-the-art architectures

Table 3
Summary of the ablation studies.

Experiments	Accuracy	Precision	Recall	F1-Score
	in (%)	in (%)	in (%)	in (%)
Swin Transformer	82.28	82.16	82.28	82.06
GSDFP	86.82	83.29	86.82	84.9
Proposed network without SA	89.92	90.35	89.89	90.12
Proposed network without Focal Loss	91.69	91.82	91.69	91.65
Proposed network with Focal Loss and SA	98.19	98.39	98.19	98.21

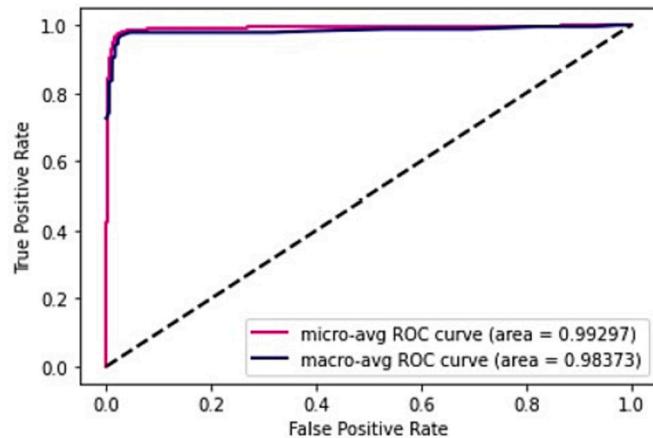


Fig. 15. ROC for the proposed network.

(90.67% - 95.88%) in terms of accuracy. It achieved highest results across other metrics such as precision (98.39%), recall (98.19%), F1-score (98.21%), and AUC (98.72%). However, it is worth noting that the proposed network had a higher number of parameters (88,389,226) compared to some of the other models such as MobileNet-V2 (2,232,839) and EfficientNet-B0 (4,016,515), which achieved accuracies of 92.07% and 94.86%, respectively. This highlights the tradeoff

between model complexity and performance, where the proposed network's higher parameter count could result in a longer training time and computational cost. Therefore, a tradeoff between the number of trainable parameters and the achieved metrics needs to be considered when selecting the appropriate model for a specific application. Overall, the results demonstrate the effectiveness of the proposed dual-track deep fusion network in improving the accuracy of citrus disease classification.

The proposed approach uses the Swin-GSDFP model. All the experiments with the model were executed on an AWS EC2 instance using PyTorch and a 24GB Nvidia A10G Tensor Core GPU. The computing environment for training consisted of an Ubuntu 20.04 operating system, 4 AMD vCPUs, and 16GB RAM. The model was trained using the Adam gradient descent optimization algorithm, with focal loss chosen as the loss function to minimize the effect of class imbalance in the training dataset. The Swin-GSDFP model has 88,389,226 trainable parameters, and it took 1.30 min to complete one epoch during the training of the model. The total training time for 100 epochs was 132 min. In conclusion, while the proposed study has achieved high accuracies, it comes at the cost of high computational complexity and long processing time for training. The choice of model should be made based on the specific needs of the application, balancing the required accuracy with the available computational resources. Further research can be done to explore more efficient models that can achieve high accuracies while reducing computational complexity and processing time.

4.7. Limitations and future work

The following are some of the limitations of the proposed network. It also proposes the scope for further research.

- The dataset used in this work is limited in size, with only ten images for the "Anthracnose" class, including two test images. This poses a risk of overfitting and poor generalization to new data. The focal loss function was applied in this work to mitigate the effects of class imbalance, ensuring that the minority class is not underweighted during training. However, in the future, there is a need to acquire more real-time images for this class for enhanced results.
- One potential direction for future work in this research article is to explore the utilization of eXplainable Artificial Intelligence (XAI)

Table 4
Performance analysis of the proposed work with current research works using Citrus Disease Image Gallery Dataset.

S No.	Source	Dataset	Classes	Methods	Accuracy in (%)	Precision in (%)	Recall in (%)
1	Elaraby et al., 2022	Citrus Disease Image Gallery, Plant Village, Citrus Plant,	Black Spot, Canker, Greening, Scab, Melanose, Healthy	AlexNet	94.3	94.1	93.9
2	Safdar et al., 2019	Citrus Disease Image Gallery, Plant Village	Alternaria, Anthracnose, Black Spot, Greening, Scab, Healthy	M-SVM	95.5	95.64	95.42
3	Proposed Network	Citrus Disease Image Gallery, Citrus Plant,	Anthracnose, Black Spot, Canker, Scab, Greening, Melanose, Healthy	Swin-GSDFP	98.19	98.39	98.19

Table 5
Performance analysis of the proposed work with current research works using Citrus Plant Dataset.

S No.	Source	Dataset	Classes	Methods	Accuracy in (%)	Precision in (%)	Recall in (%)
1	Tiwari et al., 2021	Citrus Plant	Black Spot, Canker, Greening, Healthy	Dense CNN	85.55	85.705	85.78
2	Anwar and Anwar, 2022	Citrus Plant, Plant Village	Black Spot, Canker, Greening, Melanose, Healthy	DenseNet-121	92	94	95
3	Elaraby et al., 2022	Citrus Plant, Citrus Disease Image Gallery, Plant Village	Black Spot, Canker, Greening, Scab, Melanose, Healthy	AlexNet	94.3	94.1	93.9
4	Janarthan et al., 2020	Citrus Plant	Black spot, Canker, Greening, Healthy	Custom-CNN	95.04	95.47	95.46
5	Proposed Network	Citrus Plant, Citrus Disease Image Gallery	Anthracnose, Black Spot, Canker, Scab, Greening, Melanose, Healthy	Swin-GSDFP	98.19	98.39	98.19

Table 6

Comparison with state-of-the-art architectures.

Model	Number of parameters	Accuracy	Precision	Recall	F1-Score	AUC
		in (%)				
VGG-16	134,289,223	90.67	91.37	90.37	90.61	94.05
AlexNet	57,032,519	95.88	96.08	95.80	95.79	97.41
MobileNet-v2	2,232,839	92.07	91.22	92.07	91.20	95.36
EfficientNet-B0	4,016,515	94.86	93.72	94.72	93.96	98.00
DenseNet-201	18,106,375	90.93	91.66	90.93	90.81	96.25
Proposed Network	88,389,226	98.19	98.39	98.19	98.21	98.72

techniques to elucidate the interpretation of the global and local features learned by the key components of the proposed network. This would allow for a better understanding of how the model makes its predictions, and it could also be used to improve the interpretability of the model.

- Further research in this area can aim to refine segmentation algorithms to more accurately outline disease markers in plant images. Additionally, the proposed framework can be extended to other types of plants to enable real-time computer-aided disease classification.

5. Conclusion

Citrus plant diseases are a major threat to the global citrus industry, causing significant economic losses. Traditional methods for detecting citrus diseases are time-consuming, expensive, and not scalable. Deep learning has emerged as a promising alternative for automated disease classification, as it can extract and learn complex features from images more efficiently than traditional methods. In this research, we propose a novel dual-branch deep feature fusion network for the classification of citrus plant diseases. The proposed network consists of two parallel branches: a global branch that uses a Swin transformer to extract global features, and a local branch that uses a GSDFP module to extract local features. The fused features from the two branches are then passed through an SA block to establish contextual relationships between them. Finally, a focal loss function is used to optimize the classification results.

The proposed network has several advantages over existing methods. First, it is able to extract both global and local features, which allows it to capture both the overall structure of the image and the fine-grained details. Second, the SA block allows the network to establish contextual relationships between the global and local features, which further improves the classification accuracy. Finally, the focal loss function helps to address the problem of class imbalance, which is a common problem in disease classification datasets. The proposed network was evaluated on two publicly available datasets: the Citrus Plant Dataset and the Citrus Disease Image Gallery Dataset. The results show that the proposed network achieves a classification accuracy of 98.19%, which outperforms the existing methods. The proposed network was also compared with several other state-of-the-art models, and it was found to notably outperform them.

Data availability

Data will be made available on request.

References

- Adeem, G., Rehman, Saif Ur, Ahmad, S., 2022. Classification of citrus canker and black spot diseases using a deep learning based approach. In: VFAST Transactions on Software Engineering, 10, pp. 185–197. Issue 2.
- Aidoo, O.F., Souza, P.G., da Silva, R.S., Júnior, P.A., Picanço, M.C., Osei-Owusu, J., Sétamou, M., Ekesi, S., Borgemeister, C., 2022. A machine learning algorithm-based approach (Maxent) for predicting invasive potential of *Trioza erytreae* on a global scale. Eco. Inform. 71, 101792 <https://doi.org/10.1016/j.ecoinf.2022.101792>.
- Ali, H., Lali, M.I., Nawaz, M.Z., Sharif, M., Saleem, B.A., 2017. Symptom based automated detection of citrus diseases using color histogram and textural descriptors. In: Computers and Electronics in Agriculture, vol. 138. Elsevier BV, pp. 92–104. <https://doi.org/10.1016/j.compag.2017.04.008>.
- Anwar, T., Anwar, H., 2022. Citrus plant disease identification using deep learning with multiple transfer learning approaches. Pakistan J. Eng. Technol. 3 (2), 34–38. <https://doi.org/10.51846/vol3iss2pp34-38>.
- Appiah, A.Y., Zhang, X., Ayawli, B.B.K., Kyeremeh, F., 2019. Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis. In: IEEE Access, vol. 7. Institute of Electrical and Electronics Engineers (IEEE), pp. 30089–30101. <https://doi.org/10.1109/access.2019.2902949>.
- Atila, Ü., Uçar, M., Akyol, K., Uçar, E., 2021. Plant leaf disease classification using EfficientNet deep learning model. Eco. Inform. 61, 101182 <https://doi.org/10.1016/j.ecoinf.2020.101182>.
- Bashar, A. Dr, 2019. Survey on evolving deep learning neural network architectures. In: December 2019, 2019, pp. 73–82. Issue 2. Inventive Research Organization. 10.365 48/jaicn.2019.2.003.
- Çetiner, H., 2022. Citrus disease detection and classification using based on convolution deep neural network. In: Microprocessors and Microsystems, vol. 95. Elsevier BV, p. 104687. <https://doi.org/10.1016/j.micpro.2022.104687>.
- Dananjayan, S., Tang, Y., Zhuang, J., Hou, C., Luo, S., 2022. Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. In: Computers and Electronics in Agriculture, vol. 193. Elsevier BV, p. 106658. <https://doi.org/10.1016/j.compag.2021.106658>.
- Elaraby, A., Hamdy, W., Alanazi, S., 2022. Classification of citrus diseases using optimization deep learning approach. In: Koundal, D. (Ed.), Computational Intelligence and Neuroscience, vol. 2022. Hindawi Limited, pp. 1–10. <https://doi.org/10.1155/2022/9153207>.
- Febrinato, F.G., Dewi, C., Triwiratno, A., 2019. The implementation of K-means algorithm as image segmenting method in identifying the citrus leaves disease. In: IOP Conference Series: Earth and Environmental Science, vol. 243. IOP Publishing, p. 012024. <https://doi.org/10.1088/1755-1315/243/1/012024>.
- Ferreira, J.R.M., Sierra-Garcia, I.N., Guieu, S., Silva, A.M.S., da Silva, R.N., Cunha, N., 2021, October 19. Photodynamic control of citrus crop diseases. World J. Microbiol. Biotechnol. 37 (12) <https://doi.org/10.1007/s11274-021-03171-7>.
- Ghiasi, G., Lin, T.Y., Le, Q.V., Kaiming, H., 2018. NAS-FPN: learning scalable feature pyramid architecture for object detection. In: CVPR, pp. 7036–7045.
- Image Gallery, 2023. Citrus Diseases - idtools.org. Retrieved February 12, 2023, from. <http://www.idtools.org/id/citrus/diseases/gallery.php>.
- Indra, E., Jayanthi, P.R., Venugadapathiraj, R., Balamanigandan, M., Saranya, D., Priyadarshikadevi, T., 2021. An efficient computer aided diagnosis model for citrus disease detection. In: Annals of the Romanian Society for Cell Biology, 1872–1880. Retrieved from. <https://www.annualsofrscb.ro/index.php/journal/article/view/304>.
- Janarthan, S., Thusethan, S., Rajasegarar, S., Lyu, Q., Zheng, Y., Yearwood, J., 2020. Deep metric learning based citrus disease classification with sparse data. In: IEEE Access, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 162588–162600. <https://doi.org/10.1109/access.2020.3021487>.
- Kaur, B., Sharma, T., Goyal, B., Dogra, A., 2020. A genetic algorithm based feature optimization method for citrus HLB disease detection using machine learning. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE. <https://doi.org/10.1109/icssit48917.2020.9214107>.
- Kaya, Y., Gürsoy, E., 2023. A novel multi-head CNN design to identify plant diseases using the fusion of RGB images. Eco. Inform. 75, 101998 <https://doi.org/10.1016/j.ecoinf.2023.101998>.
- Khan, E., Rehman, M.Z.U., Ahmed, F., Khan, M.A., 2021. Classification of diseases in citrus fruits using SqueezeNet. In: 2021 International Conference on Applied and Engineering Mathematics (ICAEM). IEEE. <https://doi.org/10.1109/icaem53552.2021.9547133>.
- Khattak, A., Asghar, M.U., Batool, U., Asghar, M.Z., Ullah, H., Al-Rakhami, M., Gumaei, A., 2021. Automatic detection of Citrus fruit and leaves diseases using deep neural network model. In: IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers (IEEE), pp. 112942–112954. <https://doi.org/10.1109/access.2021.3096895>.
- Li, X., Li, S., 2022. Transformer help CNN see better: a lightweight hybrid apple disease identification model based on transformers. Agriculture 12 (6), 884. <https://doi.org/10.3390/agriculture12060884>.
- Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: CVPR, pp. 2117–2125.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 1 Feb. 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42 (2), 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826>.
- Matheyambath, A.C., Padmanabhan, P., Paliyath, G., 2016. Citrus fruits. In: Encyclopedia of Food and Health. Elsevier, pp. 136–140. <https://doi.org/10.1016/b978-0-12-384947-2.00165-3>.

- Mzoughi, O., Yahiaoui, I., 2023. Deep learning-based segmentation for disease identification. *Eco. Inform.* 75, 102000 <https://doi.org/10.1016/j.ecoinf.2023.102000>.
- Prabu, M., Chelliah, B.J., 2022, August 17. An intelligent approach using boosted support vector machine based arithmetic optimization algorithm for accurate detection of plant leaf disease. *Pattern. Anal. Applic.* 26 (1), 367–379. <https://doi.org/10.1007/s10044-022-01086-z>.
- Rauf, H.T., 2019, May 28. A citrus Fruits and Leaves Dataset for Detection and Classification of citrus Diseases through Machine Learning. Mendeley Data. Retrieved February 12, 2023, from. <https://data.mendeley.com/datasets/3f83gxm57/2>.
- Safdar, A., Khan, M.A., Shah, J.H., Sharif, M., Saba, T., Rehman, A., Javed, K., Khan, J.A., 2019. Intelligent microscopic approach for identification and recognition of citrus deformities. In: *Microscopy Research and Technique*, 82. Wiley, pp. 1542–1556. <https://doi.org/10.1002/jemt.23320>. Issue 9.
- Saini, A.K., Bhatnagar, R., Srivastava, D.K., 2021a. AI based automatic detection of citrus fruit and leaves diseases using deep neural network model. *J. Discret. Math. Sci. Cryptogr.* 24 (8), 2181–2193. Informa UK Limited. <https://doi.org/10.1080/09720529.2021.2011095>.
- Saini, A.K., Bhatnagar, R., Srivastava, D.K., 2021b. Citrus fruits diseases detection and classification using transfer learning. In: Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence. DSMLAI ‘21: International Conference on Data Science, Machine Learning and Artificial Intelligence. ACM. <https://doi.org/10.1145/348424.3484893>.
- Saini, R.K., Ranjit, A., Sharma, K., Prasad, P., Shang, X., Gowda, K.G.M., Keum, Y.S., 2022, January 26. Bioactive compounds of Citrus fruits: a review of composition and health benefits of carotenoids, flavonoids, limonoids, and terpenes. *Antioxidants* 11 (2), 239. <https://doi.org/10.3390/antiox11020239>.
- Sampathkumar, S., Rajeswari, R., 2020a. An automated crop and plant disease identification scheme using cognitive fuzzy C-means algorithm. *IETE J. Res.* 68 (5), 3786–3797. Informa UK Limited. <https://doi.org/10.1080/03772063.2020.1780163>.
- Sampathkumar, S., Rajeswari, R., 2020b. An automated crop and plant disease identification scheme using cognitive fuzzy C-means algorithm. *IETE J. Res.* 68 (5), 3786–3797. Informa UK Limited. <https://doi.org/10.1080/03772063.2020.1780163>.
- Sarker, I.H., 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. In: *SN Computer Science*, vol. 2. Springer Science and Business Media LLC. <https://doi.org/10.1007/s42979-021-00815-1>. Issue 6.
- Senthilkumar, C., Kamarasan, M., 2019. Optimal segmentation with back-propagation neural network (BPNN) based citrus leaf disease diagnosis. In: 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE. <https://doi.org/10.1109/icssit46314.2019.8987749>.
- Senthilkumar, C., Kamarasan, M., 2020. An optimal weighted segmentation with hough transform based feature extraction and classification model for citrus disease. In: 2020 International Conference on Inventive Computation Technologies (ICICT). IEEE. <https://doi.org/10.1109/icict48043.2020.9112530>.
- Senthilkumar, C., Kamarasan, M., 2021. An efficient citrus detection and classification using deep learning based inception Resnet V2 model. In: 2021 Turkish Journal of Computer and Mathematics Education, 12. Proquest, pp. 2283–2296. Issue 12.
- Song, C., Wang, C., Yang, Y., 2020. Automatic detection and image recognition of precision agriculture for citrus diseases. In: 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE). IEEE. <https://doi.org/10.1109/ecice50847.2020.9301932>.
- Sudharshan Duth, P., Bhat, S.G., 2022. Disease classification in citrus leaf using deep learning. In: 2022 IEEE International Conference on Data Science and Information System (ICDSIS). IEEE. <https://doi.org/10.1109/icdsis55133.2022.9915847>.
- Sutaji, D., Yildiz, O., 2022. LEMOXINET: lite ensemble mobilenetv2 and XCEPTION models to predict plant disease. *Eco. Inform.* 70, 101698 <https://doi.org/10.1016/j.ecoinf.2022.101698>.
- Swin Transformer, 2022, July 4. Hierarchical Vision Transformer using Shifted Windows. Committed Towards Better Future. Retrieved February 12, 2023, from. <https://amaarora.github.io/2022/07/04/swintransformerv1.html>.
- Syed-Ab-Rahman, S.F., Hesamian, M.H., Prasad, M., 2021. Citrus disease detection and classification using end-to-end anchor-based deep learning model. In: *Applied Intelligence*, 52. Springer Science and Business Media LLC, pp. 927–938. <https://doi.org/10.1007/s10489-021-02452-w>. Issue 1.
- Tiwari, V., Joshi, R.C., Dutta, M.K., 2021. Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. *Eco. Inform.* 63, 101289 <https://doi.org/10.1016/j.ecoinf.2021.101289>.
- Wang, P., Fan, E., Wang, P., 2021. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. In: *Pattern Recognition Letters*, vol. 141. Elsevier BV, pp. 61–67. <https://doi.org/10.1016/j.patrec.2020.07.042>.
- Wei, J., Chu, X., Sun, X., Xu, K., Deng, H., Chen, J., Wei, Z., Lei, M., 2019. Machine learning in materials science. In: *InfoMat*, 1. Wiley, pp. 338–358. <https://doi.org/10.1002/inf2.12028>. Issue 3.
- Wu, C.-M., Chen, L.-I., 2021. Improved deep learning in citrus canker recognition system based on FPGA. In: 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). IEEE. <https://doi.org/10.1109/icce-tw52618.2021.9603192>.
- Yasmeen, U., Attique Khan, M., Tariq, U., Ali Khan, J., Asfand, E., Yar, M., Avais Hanif, C. H., Mey, S., Nam, Y., 2022. Citrus diseases recognition using deep improved genetic algorithm. In: *Computers, Materials & Continua*, 71. Computers, Materials and Continua (Tech Science Press), pp. 3667–3684. <https://doi.org/10.32604/cmc.2022.022264>. Issue 2.
- Zeng, Q., Ma, X., Cheng, B., Zhou, E., Pang, W., 2020. GANs-based data augmentation for citrus disease severity detection using deep learning. In: *IEEE Access*, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 172882–172891. <https://doi.org/10.1109/access.2020.3025196>.
- Zhang, Q.-L., Yang, Y.-B., 2021. Sa-net: Shuffle attention for deep convolutional neural networks. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). <https://doi.org/10.1109/icassp39728.2021.9414568>.
- Zhang, X., Xun, Y., Chen, Y., 2022. Automated identification of citrus diseases in orchards using deep learning. In: *Biosystems Engineering*, vol. 223. Elsevier BV, pp. 249–258. <https://doi.org/10.1016/j.biosystemseng.2022.09.006>.
- Zhong, G., Nicolosi, E., 2020. Citrus origin, diffusion, and economic importance. In: *Compendium of Plant Genomes*. Springer International Publishing, pp. 5–21. https://doi.org/10.1007/978-3-030-15308-3_2.
- Zia Ur Rehman, M., Ahmed, F., Attique Khan, M., Tariq, U., Shaukat Jamal, S., Ahmad, J., Hussain, I., 2022. Classification of citrus plant diseases using deep transfer learning. *Comp. Mater. Continua* 70 (1), 1401–1417. Computers, Materials and Continua (Tech Science Press). [10.32604/cmc.2022.019046](https://doi.org/10.32604/cmc.2022.019046).