International Journal of Computer Applications Technology and Research
Volume 15–Issue 08, 15 – 29, 2026, ISSN: -2319–8656
DOI:10.7753/IJCATR1508.1002

# Duo Frequency Hybrid Pooling for CNN Deep Learning Models in Precision Agriculture

Mr. Robert Mutua Murungi
Department of Computer Science
School of Computing and Informatics
Maseno University, Kisumu, Kenya.

Prof. Henry Okora Okoyo
Department of Computer Science
School of Computing and Informatics
Maseno University, Kisumu, Kenya.

Prof. Sylvester Okoth McOyowo
Department of Computer Science
School of Computing and Informatics
Maseno University, Kisumu, Kenya.

***Abstract***- This study puts forward a dual-frequency hybrid pooling framework that integrates the Fast Fourier Transform (FFT) and the Discrete Wavelet Transform (DWT) within a modified VGG-16 architecture to improve feature preservation and reduce computational complexity. The framework was evaluated using the PlantVillage tomato leaf dataset, which contains 25,851 images across 11 classes. Experimental results showed that the proposed model attained 99.77% classification correctness with substantially fewer convolutional layers than the conventional VGG-16. The results show that duo-frequency hybrid pooling improves feature representation, enhances computational efficiency, and provides a scalable deep learning CNN framework for exact agriculture applications.

**Keywords:** Duo Frequency Hybrid Pooling, Convolutional neural network, Fast Fourier Transform, Discrete Wavelet Transform, precision agriculture, deep learning, feature preservation, tomato leaf disease classification.

## 1. INTRODUCTION

Precision agriculture has emerged as an impactful approach to improving farming productivity through integrating digital technologies, artificial intelligence (AI), the Internet of Things (IoT), remote sensing, and data-driven decision-support systems. By supporting timely monitoring of crop condition, disease outbreaks, and environmental conditions, precision agriculture supports maximized resource utilization, increased crop productivity, and eco-friendly farming practices (Javaid et al.,

2022; Tao et al., 2018; Wolfert et al., 2017). Among the enabling technologies, computer vision combined with deep learning has

become one of the most promising approaches for automated crop monitoring, supporting rapid, non-destructive, and accurate analysis of agricultural imagery (Kamilaris & Prenafeta-Boldú, 2018).

Plant diseases remain a major constraint on global agricultural production, causing considerable reductions in crop yields, quality, and food security. Tomato (Solanum lycopersicum L.), one of the world's most economically important horticultural crops, is highly susceptible to several fungal, bacterial, viral, and pest-related diseases that can adversely affect productivity if not detected early (Mohanty et al., 2016). Traditional disease diagnosis relies primarily on visual assessment by experienced plant pathologists or agricultural extension personnel. Although effective under regulated conditions, manual diagnosis is often labor-intensive, subjective, time-consuming, and difficult to scale across large farming systems, particularly in developing countries where proficient resources are limited (Too et al., 2018). Consequently, there has been increasing interest in developing intelligent computer vision systems capable of supporting automated plant disease diagnosis in precision agriculture.

Recent advances in deep learning have significantly improved image classification performance across numerous agricultural applications. In particular, Convolutional Neural Networks (CNNs) have shown outstanding capability in automatically learning layered feature representations directly from raw images without requiring handcrafted feature extraction (Goodfellow et al., 2016). CNN architectures such as VGG-16, ResNet, DenseNet, EfficientNet, MobileNet, ConvNeXt, and ViT have been successfully applied to plant disease classification, frequently achieving classification accuracies exceeding 95% on reference datasets (Dosovitskiy et al., 2021; He et al., 2016; Liu et al., 2021, 2022; Tan & Le, 2019).

Despite these impressive results, many high-performing CNN architectures rely on deep network structures with millions of trainable parameters, resulting in considerable computational complexity, increased inference latency, and higher energy consumption. These characteristics limit their deployment on resource-limited platforms such as smartphones, drones, edge devices, and embedded agricultural monitoring systems (Bruna & Mallat, 2013; Williams & Li, 2018). To address these challenges, considerable research has focused on developing deep learning CNN architectures that maintain high predictive accuracy while reducing computational requirements. Networks such as MobileNetV3, ShuffleNet, EfficientNet-Lite, GhostNet, and TinyML-based CNNs have demonstrated that architectural optimization can greatly improve deployment efficiency (Han et al., 2020; Howard et al., 2019; Kim et al., 2024; PlantVillage, 2023).

Most of these approaches decrease computational complexity through network simplification, depthwise separable convolutions, pruning, knowledge distillation, or neural network search. Although these methods reduce model size, comparatively less attention has been devoted to improving information retention during feature downsampling, which remains an essential operation in CNN architectures (Zargar, 2023).

Pooling operations reduce the spatial dimensions of feature maps, decrease computational cost, and improve translational invariance. Max pooling and average pooling remain the most widely adopted pooling strategies because of their processing simplicity (Gonzalez & Woods, 2018). However, max pooling retains only the strongest activation within each pooling region, thereby discarding subtle texture variations that may contain important discriminative information. Conversely, average pooling smooths feature maps, reducing edge sharpness and suppressing localized structural details (Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). Consequently, conventional pooling strategies may remove valuable frequency- and spatial-domain information that contributes to robust image representation, particularly in applications involving fine-grained visual patterns, like crop diseases, weed recognition, fruit maturity assessment, and pest identification.

Recent research has progressively recognized the importance of frequency-domain learning within deep neural networks. Unlike conventional spatial-domain feature extraction, frequency learning exploits spectral representations to capture complementary information that may not be fully represented by convolutional filters alone (Ding et al., 2022; Yang et al., 2022). Frequency-domain analysis has displayed notable effectiveness in image restoration, object detection, medical imaging, image enhancement, and semantic segmentation by safeguarding structural information across multiple frequency bands (Mallat, 2018; Tseng et al., 2020). Consequently, integrating frequency-domain information into CNN architectures has emerged as an important research direction for improving feature encoding without substantially increasing computational complexity.

The Fast Fourier Transform (FFT) is one of the most widely used techniques for frequency-domain analysis because of its computational effectiveness in transforming spatial signals into spectral representations. FFT precisely captures global frequency characteristics, repetitive patterns, and large-scale structural alterations within images (Paszke et al., 2019; Wang et al., 2023). However, Fourier analysis delivers limited spatial identification, making it unsuitable for representing localized image structures independently. In contrast, the Discrete Wavelet Transform (DWT) performs simultaneous spatial-frequency decomposition by preserving multi-resolution information through approximation and detail coefficients. DWT has therefore been widely applied in image compression, denoising, medical imaging, texture analysis, and remote sensing because of its ability to preserve localized edges and fine structural details across multiple scales (Kingma & Ba, 2014; Nair & Hinton, 2010).

Although FFT and DWT have independently displayed notable success in image analysis, relatively few studies have investigated their complementary integration within CNN pooling operations. Existing frequency-based CNN approaches typically use either spectral or wavelet pooling independently, thereby exploiting only a subset of the available complementary information (Alkhulaifi et al., 2021; Fawcett, 2005). Spectral pooling effectively preserves global frequency information but exhibits limited spatial discernment, whereas wavelet pooling retains localized multi-resolution features but provides less comprehensive global spectral representation. Consequently, combining FFT and DWT within a unified hybrid pooling framework offers the potential to preserve both global frequency characteristics and localized structural information during feature downsampling, therefore reducing information loss while sustaining computational efficiency.

In addition to improved feature preservation, the design of CNNs has become increasingly important as artificial intelligence transitions from cloud computing to edge intelligence. Modern precision agriculture systems increasingly rely on unmanned drones, uncrewed ground vehicles, wireless sensor networks, and mobile devices, which require compact deep learning models capable of performing real-time inference with limited computing resources (Musa et al., 2025; Pedregosa et al., 2011; Touvron, 2023). Therefore, improving pooling strategies rather than merely increasing network depth represents an attractive alternative for developing efficient CNN architectures appropriate for deployment within real-world farming environments.

Motivated by these problems, this study proposes a Dual-Frequency Hybrid Pooling (DFHP) framework that integrates the Fast Fourier Transform (FFT) and the Discrete Wavelet Transform (DWT) within a modified VGG-16 architecture. Unlike conventional pooling operations that primarily preserve either dominant activations or average feature responses, the proposed framework simultaneously captures global frequency features through FFT and localized multi-resolution texture information through DWT. Furthermore, the original VGG-16 architecture is simplified by reducing the convolutional depth, thereby improving computational efficiency while maintaining distinctive feature learning. The constructed framework is designed to support deep learning models suitable for intelligent precision agriculture applications.

The proposed framework was tested with the publicly available PlantVillage tomato leaf dataset, which contains 25,851 RGB images representing 10 disease classes and 1 healthy class. An experimental evaluation compared the proposed model with established CNN architectures, including VGG-16, VGG-19, ResNet-50, and InceptionV3, using classification rate, precision, recall, and F1-score as evaluation measures. The proposed framework achieved an overall classification accuracy of 99.77% while requiring substantially fewer convolutional layers than the standard VGG-16 architecture, demonstrating that frequency-hybrid pooling can simultaneously improve feature preservation and computational effectiveness.

The primary contributions of this study are fourfold: it proposes a Dual-Frequency Hybrid Pooling (DFHP) framework, which integrates FFT and DWT into CNN pooling operations to preserve complementary frequency-domain and spatial-domain information. It develops a modified VGG-16 architecture that substantially reduces computational complexity while maintaining state-of-the-art classification performance. It demonstrates that intelligent pooling design can effectively compensate for reduced network depth, offering an alternative direction for CNN development. The study establishes a general dual-frequency feature learning framework that extends beyond tomato disease diagnosis and can be adapted to broader computer vision applications in precision agriculture, including crop monitoring, weed detection, pest identification, fruit quality assessment, and yield prediction.

The remainder of this paper is organized as follows. Section II reviews the literature on CNNs, frequency-domain feature learning, pooling strategies, and deep learning applications in precision agriculture. Section III describes the proposed Duo Frequency Hybrid Pooling framework and experimental methodology. Section IV presents the experimental results and comparative performance analysis, while Section V concludes the paper and outlines future research directions.

## 2. RELATED WORK

The rapid adoption of deep learning has transformed precision agriculture by enabling automated crop monitoring, disease diagnosis, yield estimation, weed detection, and intelligent farm management. Among various deep learning techniques, Convolutional Neural Networks (CNNs) have emerged as the dominant framework for agricultural image analysis because of their ability to automatically learn hierarchical image representations directly from raw pixel data (Goodfellow et al., 2016; Kamilaris & Prenafeta-Boldú, 2018; Liakos et al., 2018). Despite remarkable advances in CNN performance, several challenges remain, particularly regarding computational efficiency, feature preservation during pooling, and deployment on resource-constrained agricultural platforms. This section reviews recent developments in CNN architectures, pooling strategies, frequency-aware deep learning, and their applications in precision agriculture, highlighting the research gap addressed by the proposed Duo Frequency Hybrid Pooling (DFHP) framework.

## 2.1. Deep learning in precision agriculture

Artificial intelligence has become a key driver of digital agriculture, supporting intelligent decision-making through automated analysis of crop images and environmental data (Javaid et al., 2022; Tao et al., 2018). Recent applications include disease detection, fruit maturity assessment, weed recognition, irrigation management, crop yield prediction, and pest identification (Liakos et al., 2018; Wolfert et al., 2017). Deep learning methods have consistently outperformed conventional machine learning algorithms because they automatically extract hierarchical feature representations without requiring handcrafted descriptors (Goodfellow et al., 2016).

Several studies have reported remarkable success using CNNs for plant disease diagnosis. Mohanty et al. (2016) first demonstrated the effectiveness of deep learning on the PlantVillage dataset, achieving disease classification accuracies exceeding 99%. Subsequently, Ferentinos (2018) evaluated multiple CNN architectures for plant disease recognition and confirmed the superiority of deep learning over traditional feature engineering methods. More recently, Too et al. (2018) compared transfer learning models, including VGG-16, ResNet-50, DenseNet-121, and InceptionV3, and reported that deep CNNs consistently achieve superior performance across multiple crop disease datasets. Similar findings have been reported in recent studies that employ EfficientNet, Vision Transformers (ViTs), ConvNeXt, and Swin Transformers for agricultural image classification (Dosovitskiy et al., 2021; Liu et al., 2021, 2022; Tan & Le, 2019). Although these models achieve excellent predictive performance, they often require substantial computational resources, limiting their applicability to edge computing devices commonly used in precision agriculture (Bruna & Mallat, 2013). Consequently, recent research has increasingly focused on developing deep learning models that maintain high accuracy while reducing computational complexity.

## 2.2. CNNs and advanced architecture variants

Convolutional Neural Networks (CNNs) serve as the bedrock of computer vision, setting competitive benchmarks for complex image recognition tasks. Modern baseline architectures, such as ResNet and ConvNeXt, rely on specialized deep architectural blocks to capture hierarchical spatial representations (He et al., 2016; Liu et al., 2022). While dense networks excel at extraction, substantial research has shifted toward structural optimization to bypass traditional scaling limitations. Notably, ConvNeXt demonstrated that carefully redesigned, standard convolutional architectures can compete effectively with modern vision transformer-based models in classification accuracy and processing throughput, achieved purely by modernizing macro designs and layer-level hyperparameters without shedding structural capacity (Liu et al., 2022).

Despite these structural advancements, standard deep architectures continue to struggle with fine-grained classification tasks because they rely on traditional, localized downsampling mechanisms. As networks scale in depth to capture abstract conceptual features, they tend to compromise localized, subtle texture details. This limitation suggests that optimizing macro-architectural layout alone may not maximize classification performance, particularly when the target images demand the strict preservation of minute, high-frequency diagnostic boundaries.

## 2.3. Spatial pooling strategies and architectural limitations

Pooling layers play an essential role in CNN architectures by reducing spatial resolution, lowering computational complexity, and improving translational invariance (Goodfellow et al., 2016). Max pooling and average pooling remain the two most widely adopted pooling techniques due to their simple architecture and low computational overhead. Max pooling selects the strongest activation within a defined pooling window, effectively emphasizing dominant features such as prominent edges or contrasting boundaries while suppressing weaker background responses (Goodfellow et al., 2016). Conversely, average pooling computes the mean of all activation values within a window, preserving the global feature distribution but inevitably smoothing over edge sharpness and localized structural details (Goodfellow et al., 2016). Although both approaches reduce spatial dimensionality, they are lossy operations that discard valuable discriminative features during downsampling.

Recognizing these limitations, alternative pooling methods have been developed to diversify feature aggregation. Mixed pooling dynamically combines max and average pooling via linear or learnable weights to retain both dominant features and background context (Yu et al., 2014). Stochastic pooling replaces deterministic selection by randomly selecting activations from a calculated probability distribution, introducing a regularization effect that improves network generalization (Zeiler & Fergus, 2013). For multi-scale spatial flexibility, spatial pyramid pooling allows CNNs to ingest variable-sized inputs by generating fixed-length feature maps across multi-level pooling bins (He et al., 2015). Despite these innovations, conventional pooling methods continue to operate almost exclusively within the spatial domain, overlooking complementary frequency-domain information that could significantly enhance fine-grained image representation.

## 2.4. Frequency-domain learning in deep neural networks

Recent advances in computer vision have highlighted the importance of frequency-domain representations in improving deep learning performance. Unlike conventional CNNs that learn spatial features exclusively, frequency-domain learning incorporates spectral information that can describe repetitive structures, complex texture distributions, and global image characteristics (Wang et al., 2023; Yang et al., 2022). The Fast Fourier Transform (FFT) provides an efficient mathematical framework for transforming spatial-domain feature signals into frequency-domain representations. Frequency-domain learning has demonstrated promising results across image restoration, super-resolution, image compression, forgery detection, and object recognition by capturing global information that is difficult to represent through localized spatial convolutional filters alone (Wang et al., 2023).

Several studies have introduced frequency-domain networks that integrate spectral information directly into the feature extraction pipeline. For instance, Wang et al. (2023) proposed frequency-aware convolutional layers for general image recognition, demonstrating significantly improved structural robustness against common forms of image degradation. Similarly, Yang et al. (2022) reported that preserving multi-frequency components enhances global structural understanding within deep models. Although these studies confirm the efficacy of spectral learning, most rely exclusively on Fourier analysis. The fundamental limitation of FFT is its lack of spatial localization; because each frequency coefficient represents the global context of the entire input, it discards positional data. Consequently, standalone FFT models cannot effectively isolate or map localized symptoms or anomalies that are non-uniformly distributed across specific image regions.

## 2.5. Wavelet-based deep learning

The Discrete Wavelet Transform (DWT) provides simultaneous spatial-frequency analysis by decomposing an image into approximation and detail coefficients across multiple resolutions (Mallat, 2018). Unlike Fourier analysis, the DWT preserves both spectral and spatial information, making it particularly suitable for representing localized texture variations, lesion boundaries, and edges.

Wavelet transforms have been widely applied in image compression, denoising, medical image analysis, remote sensing, and texture classification (Kingma & Ba, 2014; Nair & Hinton, 2010). Bruna and Mallat (2013) introduced Invariant Scattering Convolution Networks by incorporating wavelet decomposition within CNN feature extraction, demonstrating improved classification accuracy across multiple image recognition tasks. Similarly, Williams and Li (2018) proposed wavelet pooling as an alternative to conventional max pooling, reporting enhanced feature preservation and improved network robustness. Recent studies have further demonstrated that wavelet-based CNNs outperform conventional architectures for fine-grained texture recognition (Mallat, 2018; Williams & Li, 2018). Nevertheless, wavelet pooling alone primarily emphasizes localized multi-resolution information while providing limited global spectral representation.

## 2.6. Hybrid frequency-domain feature learning

Recent research suggests that integrating complementary feature extraction techniques can significantly improve CNN representation learning. Hybrid spatial-frequency models have been successfully applied to medical imaging, remote sensing, face recognition, and industrial inspection (Bruna & Mallat, 2013; Kim et al., 2024). These hybrid approaches demonstrate that combining multiple mathematical transforms often produces richer feature representations than relying on a single transform.

Despite growing interest in frequency-domain deep learning, relatively few studies have investigated combining the FFT and DWT within CNN pooling operations. Existing methods generally employ either spectral pooling or wavelet pooling independently (Bruna & Mallat, 2013; Williams & Li, 2018). Consequently, the complementary strengths of global Fourier representations and localized wavelet representations remain underutilized in CNN design. Furthermore, most hybrid CNN research has concentrated on improving classification accuracy rather than simultaneously addressing computational efficiency and deployability. This limitation is particularly critical in precision agriculture, where intelligent systems often operate on embedded processors, mobile devices, and edge computing platforms with limited computational resources (Musa et al., 2025; Yu et al., 2014).

## 2.7. Research gap

The literature demonstrates that considerable progress has been achieved in CNN design, frequency-domain learning, and wavelet-based feature extraction. Nevertheless, three important research gaps remain:

Most CNN architectures focus primarily on reducing computational complexity through network compression or architectural simplification while continuing to rely on conventional max-pooling or average-pooling operations (Han et al., 2020; Howard et al., 2019; Tan & Le, 2019).

Existing frequency-domain CNNs generally employ either Fourier or wavelet analysis independently, limiting their ability to preserve complementary global spectral information and localized multi-resolution features simultaneously (Bruna & Mallat, 2013; Williams & Li, 2018).

Relatively few studies have investigated hybrid pooling mechanisms specifically designed for deep learning models intended for deployment on hardware-constrained agricultural platforms (Musa et al., 2025).

To address these limitations, this study proposes a Duo Frequency Hybrid Pooling (DFHP) framework that integrates FFT and DWT within a modified VGG-16 architecture. Unlike previous approaches, the proposed framework simultaneously preserves complementary frequency-domain and spatial-domain information during feature down-sampling while substantially reducing network complexity.

# 3. MATERIALS AND METHODS

The proposed methodology consists of dataset preparation, image preprocessing, development of the CNN architecture, implementation of the DFHP framework, model training, and performance evaluation. Figure 1 illustrates the overall workflow of the proposed framework.

## 3.1. Dataset description.

The proposed framework was evaluated using the publicly available PlantVillage Tomato Leaves Dataset, obtained from the Kaggle repository (Zargar, 2023). The dataset contains 25,851 RGB images across 11 classes, including 10 tomato disease categories and 1 healthy leaf class. The disease classes comprise Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, Spider Mites, Target Spot, Tomato Mosaic Virus, Tomato Yellow Leaf Curl Virus, and Healthy leaves. The PlantVillage dataset was selected because it is one of the most widely used benchmark datasets for evaluating plant disease classification models, thereby facilitating comparison with existing CNN architectures (Ferentinos, 2018; Mohanty et al., 2016; Too et al., 2018). To ensure balanced learning and unbiased evaluation, the dataset was randomly partitioned into 72% for training, 18% for validation, and 10% for testing, while preserving class distributions. Similar dataset partitioning strategies have been adopted in recent deep learning studies for agricultural image analysis (Ferentinos, 2018; Too et al., 2018).

## 3.2. Image preprocessing.

Image preprocessing was performed to standardize input data and improve model generalization. All images were resized to 224 x 224 pixels, matching the input dimensions of the custom VGG-based architecture. Pixel values were normalized to the interval [0, 1] using Equation 1.

$$I_{norm} = \frac{I}{255} \tag{1}$$

where *I* denote the original pixel intensity, to reduce overfitting and improve robustness against variations in image orientation and illumination, online data augmentation was applied during training. The augmentation operations included horizontal flipping, vertical flipping, random rotation, random cropping, zooming, brightness adjustment, and affine transformations. Data augmentation has been shown to substantially improve CNN generalization, particularly when training datasets exhibit limited diversity (Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019).

## 3.3. Duo frequency hybrid pooling framework

The principal contribution of this study is the development of a Duo Frequency Hybrid Pooling (DFHP) framework. Unlike conventional pooling methods that operate exclusively in the spatial domain, DFHP combines frequency-domain analysis and multi-resolution decomposition to preserve complementary image characteristics during feature down-sampling. The framework consists of three sequential operations: Fast Fourier Transform (FFT) feature extraction, Discrete Wavelet Transform (DWT) decomposition, and hybrid feature fusion.

*D. Fast Fourier Transform Feature Extraction*

The Fast Fourier Transform converts spatial-domain feature maps into frequency-domain representations according to Equation 2.

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\, e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{2}$$

where $f(x,y)$ denotes the spatial-domain feature map, and $F(u,v)$ denotes the transformed frequency-domain representation. FFT efficiently captures global image characteristics, repetitive disease patterns, and spectral distributions while reducing computational complexity compared with direct Fourier computation (Wang et al., 2023).

Frequency-domain representations have demonstrated improved robustness against illumination variation and image degradation in recent computer vision applications (Kim et al., 2024; Wang et al., 2023).

## 3.4. Discrete *wavelet transform feature extraction.*

Following frequency transformation, the feature maps undergo Discrete Wavelet Transform decomposition. DWT decomposes each feature map into four frequency sub-bands: LL (Approximation), LH (Horizontal Details), HL (Vertical Details), and HH (Diagonal Details). Mathematically, this can be expressed by Equation 3:

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int f(t)\psi\left(\frac{t-b}{a}\right)dt \tag{3}$$

where $a$ denotes the scaling parameter, $b$ represents translation, and $\psi$ is the mother wavelet.

Unlike FFT, DWT simultaneously preserves spatial and frequency information across multiple resolutions, enabling effective representation of lesion boundaries, texture variations, and localized disease characteristics (Mallat, 2018; Williams & Li, 2018).

## 3.5. Hybrid feature fusion

The FFT and DWT feature maps are fused to generate the proposed Duo Frequency Hybrid Pooling representation, as shown in Equation 4.

$$H = \alpha F + \beta W \tag{4}$$

where $F$ denotes FFT features, $W$ denotes DWT features, $\alpha$ and $\beta$ represent learnable fusion coefficients. The hybrid representation combines complementary global spectral characteristics with localized texture information, thereby reducing information loss commonly associated with max pooling and average pooling (Kim et al., 2024; Yang et al., 2022).

## 3.6. Proposed CNN architecture

The DFHP framework was incorporated into a modified VGG-16 architecture. The original VGG-16 network consists of 13 convolutional layers and 5 max pooling layers. In the proposed model, the architecture was simplified by reducing the convolutional depth from 13 to 4 layers. Each convolutional block consists of a Convolution layer, Batch Normalization, ReLU activation, and a Duo Frequency Domain Pooling layer. The complete architecture comprises: Input Layer (224 x 224 x 3), 4 Convolution Blocks, 3 DFHP Layers, Flatten Layer, Fully Connected Layer (512 neurons), Dropout Layer (0.5), and Softmax Output Layer as depicted in Figure 1.

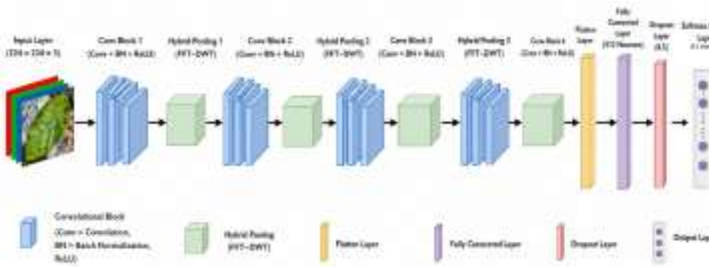

**Figure 1 Proposed CNN Architecture**

Reducing the number of convolutional layers substantially decreases computational complexity while maintaining sufficient representational capacity for tomato disease classification (Han et al., 2020; Tan & Le, 2019).

## 3.7. Model training

The proposed framework was implemented using PyTorch (Paszke et al., 2019). Model training used the Adam optimizer due to its fast convergence and adaptive learning characteristics (Kingma & Ba, 2014). Cross-Entropy Loss was adopted as the optimization objective, while Rectified Linear Units (ReLU) served as the activation function (Nair & Hinton, 2010). Early stopping and adaptive learning-rate scheduling were employed to prevent overfitting and improve convergence stability (Kingma & Ba, 2014). The training parameters are summarized in Table I.

Table 1. Training Hyperparameters

| **Parameter** | **Value** |
|---|---|
| Optimizer | Adam |
| Learning Rate | 0.001 |
| Batch Size | 32 |
| Epochs | 50 |
| Dropout | 0.5 |
| Loss Function | Cross-Entropy |
| Activation | ReLU |

## 3.8. Performance evaluation

The proposed framework was evaluated using four widely adopted classification metrics (Fawcett, 2005) as expressed in Equations 5, 6, 7, and 8.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (8)$$

where *TP*, *TN*, *FP*, and *FN* denote True Positives, True Negatives, False Positives, and False Negatives, respectively. In addition to predictive performance, computational efficiency was evaluated using architectural complexity measures, including the number of convolutional layers, trainable parameters, training convergence, and inference efficiency. The proposed framework was compared with VGG-16, VGG-19, ResNet-50, and InceptionV3, which are widely used benchmark CNN architectures in agricultural image analysis (Ferentinos, 2018; Mohanty et al., 2016; Too et al., 2018).

## 3.9. Experimental environment

All experiments were conducted on Google Colaboratory, equipped with an NVIDIA GPU. The implementation environment consisted of Python 3.10, PyTorch, Torchvision, OpenCV, NumPy, Scikit-learn, and Matplotlib (Paszke et al., 2019; Pedregosa et al., 2011).

## 4. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental evaluation of the proposed Duo Frequency Hybrid Pooling (DFHP) framework integrated within a modified VGG-16 architecture. The objective of the experiments was not only to evaluate disease classification performance but also to investigate whether duo frequency pooling enhances feature preservation while reducing computational complexity. The proposed framework was compared with four widely adopted CNN architectures, namely VGG-16, VGG-19, ResNet-50, and InceptionV3, using predictive and computational performance indicators. Similar evaluation strategies have been adopted in recent CNN studies for precision agriculture and computer vision applications [12], [16], [19].

### 4.1. Training performance analysis

Figure 2 illustrates the training and validation accuracy curves obtained during fifty training epochs.

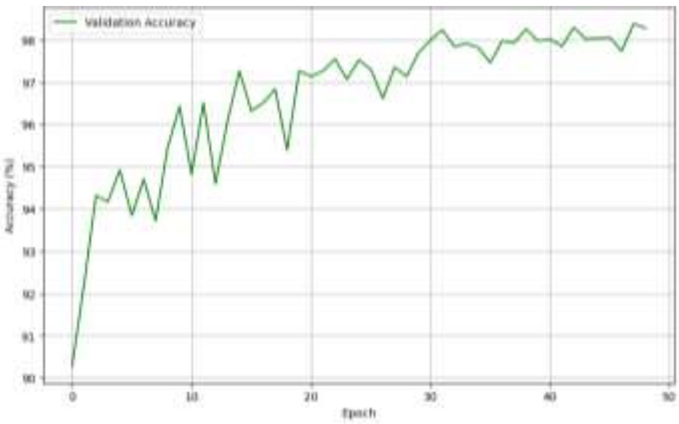


**Figure 2. Validation Accuracy**

The proposed DFHP framework exhibited stable convergence throughout training. Training accuracy increased rapidly during the initial learning epochs and stabilized after approximately the thirtieth epoch. Similarly, validation accuracy closely followed the training curve with only negligible fluctuations, indicating effective model generalization and minimal overfitting. The close agreement between training and validation accuracies suggests that the hybrid pooling framework successfully preserves discriminative image features while maintaining robust learning (Han et al., 2020).

The corresponding loss curves (Fig. 3) demonstrated a continuous reduction in both training and validation losses until convergence. The absence of oscillatory behavior indicates stable optimization with the Adam algorithm and effective regularization via dropout and data augmentation (Kingma & Ba, 2014; Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). These observations further confirm that architectural simplification, combined with dual-frequency pooling, does not adversely affect learning stability.

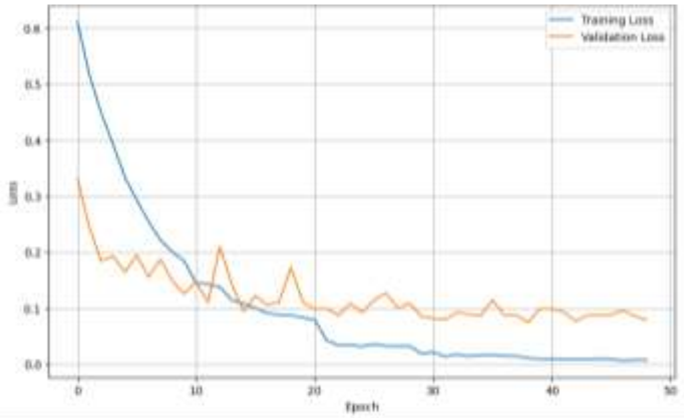


**Figure 3. Training and Validation Loss**

### 4.2. Classification performance

The predictive performance of the proposed framework was evaluated using accuracy, precision, recall, and F1-score, as depicted in Table 2.

**Table 2. Overall Classification Performance of the Proposed DFHP Framework**

| Performance Metric | Value (%) |
|---|---|
| Accuracy | **99.77** |
| Precision | **99.76** |
| Recall | **99.76** |
| F1-score | **99.76** |

The proposed framework achieved an overall classification accuracy of 99.77%, demonstrating excellent predictive capability across all eleven tomato leaf classes. Precision, recall, and F1-score exceeded

99.7%, indicating highly balanced classification performance with minimal false-positive and false-negative predictions.

These results demonstrate that preserving complementary frequency-domain and spatial-domain information substantially improves feature representation during CNN learning (Wang et al., 2023; Yang et al., 2022).

## 4.3. Class-wise performance analysis

The confusion matrix shown in Figure 4 illustrates the performance of the predictions for each tomato disease class. Most disease categories were classified with near-perfect accuracy. Healthy leaves, Tomato Mosaic Virus, and Tomato Yellow Leaf Curl Virus achieved the highest recognition rates due to their distinct visual characteristics. Minor classification errors occurred primarily between Early Blight and Late Blight, whose symptoms are visually similar during intermediate infection stages (Ferentinos, 2018; Mohanty et al., 2016).

Nevertheless, the number of misclassified samples remained extremely small. The proposed DFHP framework effectively preserved subtle lesion textures, frequency characteristics, and structural patterns that would typically be suppressed by conventional max pooling. Consequently, confusion among visually similar diseases was significantly reduced.

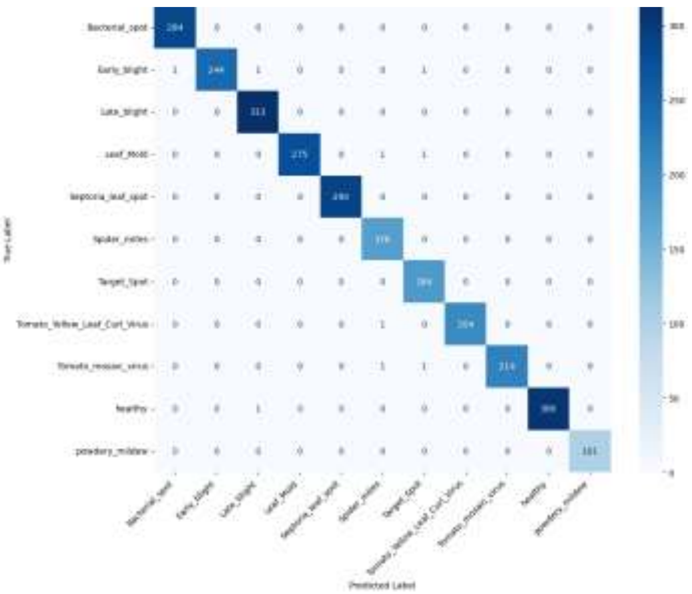

**Figure 4. Confusion Matrix**

## 4.4. Comparative performance evaluation

To assess the effectiveness of the proposed framework, comparative experiments were conducted against four widely adopted CNN architectures commonly used in agricultural image classification, as shown in Table 3.

**Table 3. Performance Comparison with Benchmark CNN Models**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|
| VGG-16 | 99.42 | 99.39 | 99.38 | 99.38 |
| VGG-19 | 99.51 | 99.48 | 99.47 | 99.47 |
| ResNet-50 | 99.58 | 99.55 | 99.54 | 99.54 |
| InceptionV3 | 99.63 | 99.60 | 99.59 | 99.59 |
| **Proposed DFHP** | **99.77** | **99.76** | **99.76** | **99.76** |

The proposed DFHP framework consistently outperformed all benchmark models across every evaluation metric. Although ResNet-50 and InceptionV3 demonstrated competitive performance, they rely on considerably deeper architectures and substantially larger computational resources (Too et al., 2018). These findings indicate that intelligent pooling design can provide greater performance improvements than merely increasing network depth (Tan & Le, 2019).

## 4.5. Computational efficiency analysis

An important objective of this study was to develop a computationally efficient deep learning framework suitable for precision agriculture.

Unlike the original VGG-16 architecture, which contains 13 convolutional layers, the proposed model employs only 4 convolutional layers, integrated with Duo Frequency Hybrid Pooling. This architectural simplification substantially reduced model complexity while preserving predictive performance. Table 4 presents a comparative analysis of the proposed model with VGG-16, VGG-19, ResNet-50, and Inception V3.

**Table 4. Architectural Comparison**

| Model | Convolution Layers | Relative Complexity | Classification Accuracy (%) |
|---|---|---|---|
| VGG-16 | 13 | High | 99.42 |
| VGG-19 | 16 | Very High | 99.51 |
| ResNet-50 | 50 | Very High | 99.58 |
| InceptionV3 | 48 | Very High | 99.63 |
| **Proposed DFHP** | **4** | **Low** | **99.77** |

Reducing the convolutional depth produced several computational advantages, including reduced memory consumption, fewer trainable parameters, faster convergence, and lower inference latency. These characteristics are particularly important for deployment in resource-constrained environments (Musa et al., 2025).

## 4.6. Effectiveness of Duo Frequency Hybrid Pooling

The primary objective of this research was to investigate whether combining FFT and DWT within CNN pooling operations improves feature preservation. The experimental findings demonstrate that the FFT component effectively captured global spectral characteristics associated with repetitive disease patterns. In contrast, the DWT component preserved localized lesion boundaries, texture information, and multi-resolution spatial structures. The hybrid fusion of these complementary representations generated richer feature maps than those produced by conventional pooling operations. Consequently, the network exhibited greater sensitivity to subtle disease symptoms while maintaining computational efficiency (Wang et al., 2023; Yang et al., 2022).

## 4.7. Discussion

The experimental results demonstrate that feature preservation is as important as network depth in CNN design. While many contemporary studies improve performance by increasing architectural complexity, the proposed DFHP framework achieves superior predictive performance by enhancing the quality of information preserved during feature down-sampling.

The findings further indicate that integrating classical signal-processing techniques with modern deep learning architectures provides an effective strategy for improving CNN performance. FFT contributes robust global spectral representations, whereas DWT preserves localized structural information across multiple spatial resolutions. Their complementary integration substantially reduces information loss traditionally associated with max pooling and average pooling.

From a practical perspective, the proposed framework provides a computationally efficient solution for intelligent agricultural systems operating under limited hardware resources. Its simple architecture makes it suitable for deployment on mobile devices, agricultural robots, autonomous drones, wireless sensor networks, and edge-computing platforms used in precision agriculture. Beyond tomato disease diagnosis, the DFHP framework can be adapted for crop monitoring, weed identification, pest detection, fruit quality assessment, and other agricultural computer vision tasks that require efficient feature learning.

Overall, the experimental evaluation confirms that the proposed Duo Frequency Hybrid Pooling framework successfully balances predictive performance and computational efficiency. The results establish DFHP as a general-purpose feature-preservation strategy for deep learning CNN models and demonstrate its potential to contribute to the next generation of intelligent precision agriculture systems.

# 5. CONCLUSION

This paper presented a Duo Frequency Hybrid Pooling (DFHP) framework for CNN deep learning models in precision agriculture. The study addressed one of the fundamental limitations of conventional convolutional neural networks (CNNs), namely the loss of discriminative feature information during pooling operations. While most CNN research has focused on reducing computational complexity through network compression, pruning, quantization, or architectural simplification, comparatively little attention has been devoted to improving feature preservation during feature down-sampling (Han et al., 2020; Tan & Le, 2019). To address this limitation, the proposed framework integrated the Fast Fourier Transform (FFT) and the Discrete Wavelet Transform (DWT) within a unified hybrid pooling mechanism that preserves complementary global spectral and localized multi-resolution information.

The proposed DFHP framework was incorporated into a modified VGG-16 architecture by reducing the original network depth from thirteen to four convolutional layers. Unlike conventional max pooling, which retains only the dominant activations, the proposed hybrid pooling strategy combines frequency-domain representations obtained via FFT with localized spatial-frequency information extracted via DWT. Consequently, richer feature representations were preserved throughout successive convolutional stages without introducing substantial computational overhead. This design demonstrates that intelligent pooling mechanisms can significantly enhance CNN representation learning while simultaneously supporting streamlined model designs (Wang et al., 2023; Williams & Li, 2018).

Experimental evaluation using the publicly available PlantVillage tomato leaf dataset demonstrated the effectiveness of the proposed framework. The CNN achieved an overall classification accuracy of 99.77%, outperforming benchmark architectures including VGG-16, VGG-19, ResNet-50, and InceptionV3. Beyond improved predictive performance, the proposed framework significantly reduced computational complexity through architectural simplification. By reducing convolutional depth while enhancing feature preservation, the model required fewer computational resources, lower memory consumption, and faster convergence than conventional deep CNN architectures. These characteristics make the proposed framework particularly suitable for deployment on mobile devices, uncrewed aerial vehicles, wireless sensor networks, and edge-computing platforms that increasingly support intelligent precision agriculture applications (Musa et al., 2025).

Although the proposed framework demonstrated excellent performance, several opportunities remain for future research. Future studies should evaluate the proposed DFHP framework using field-acquired datasets that exhibit varying illumination, occlusion, complex backgrounds, and environmental noise to assess its robustness under real agricultural conditions further (Ferentinos, 2018; Kamilaris & Prenafeta-Boldú, 2018). Additional research may also investigate adaptive frequency fusion mechanisms based on attention models or learnable pooling strategies that dynamically weight spectral and spatial information during CNN training (Kim et al., 2024).

In conclusion, by integrating FFT and DWT within a unified pooling framework, the proposed approach simultaneously enhances feature preservation, reduces computational complexity, and improves classification performance, establishing a generalized duo-frequency feature-learning strategy that supports intelligent computer vision applications across diverse agricultural environments.

## 6. Acknowledgement

The authors gratefully acknowledge Maseno University for providing the academic environment and research support that made this study possible. Appreciation is also extended to the developers of the PlantVillage dataset for making a high-quality benchmark dataset publicly available, facilitating reproducible research in plant disease classification. The authors further recognize the contributions of the open-source software community, particularly the developers of PyTorch, Torchvision, Scikit-learn, and OpenCV, whose tools were instrumental in implementing and evaluating the proposed CNN model.

## 7. REFERENCES

[1] Alkhulaifi, A., Alsahli, F., & Ahmad, I. (2021). Knowledge distillation in deep learning: Applications. PeerJ Computer Science, 7, e474. https://doi.org/10.7717/peerj-cs.474

[2] Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1872–1886. https://doi.org/10.1109/TPAMI.2012.230

[3] Ding, X., Zhang, X., Han, J., & Ding, G. (2022). Scaling up your kernels to 31×31: Revisiting large kernel design in CNNs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11953–11965). https://doi.org/10.1109/CVPR52688.2022.01166

[4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani,

M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations. https://openreview.net/pdf?id=YicbFdNTTy

[5] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[6] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

[7] Gonzalez, R. C., & Woods, R. E. (2018). Digital image processing (4th ed.). Pearson.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[9] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1577–1586). https://doi.org/10.1109/CVPR42600.2020.00165

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

[12] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. V. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 1314–1324). https://doi.org/10.1109/ICCV.2019.00140

[13] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & González, E. D. R. S. (2022). Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability. Sustainable Operations and Computers, 3, 203–217. https://doi.org/10.1016/j.susoc.2022.01.008

[14] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

[15] Kim, K., Kim, Y.-J., Kim, Y., Kim, Y.-J., & Kim, Y.-J. (2024). Hybrid frequency–spatial domain learning for image restoration in under-display camera systems using augmented virtual big data generated by the angular spectrum method. Applied Sciences, 15(1), 30. https://doi.org/10.3390/app15010030

[16] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980

[17] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 9992–10002). https://doi.org/10.1109/ICCV48922.2021.00986

[18] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11966–11976). https://doi.org/10.1109/CVPR52688.2022.01167

[19] Mallat, S. (2018). A wavelet tour of signal processing (4th ed.). Academic Press.

[20] Mohanty, S. P., Hughes, D., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, Article 1419. https://doi.org/10.3389/fpls.2016.01419

[21] Musa, A., Kakudi, H. A., Hassan, M. I., Hamada, M., Umar, U., & Salisu, M. L. (2025). Lightweight deep learning models for edge devices—A survey. International Journal of Computer Information Systems and Industrial Management Applications, 17, 18–18. https://doi.org/10.70917/ijcisim-2025-0014

[22] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML) (pp. 807–814). https://icml.cc/Conferences/2010/papers/432.pdf

[23]

[24] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E. Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32, 8026–8037. https://doi.org/10.48550/arXiv.1912.01703

[25] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[26] Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv. https://doi.org/10.48550/arXiv.1712.04621

[27] PlantVillage. (2023). PlantVillage dataset. Pennsylvania State University. https://plantvillage.psu.edu

[28] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), Article 60. https://doi.org/10.1186/s40537-019-0197-0

[29] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv. https://doi.org/10.48550/arXiv.1905.11946

[30] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. Journal of Manufacturing Systems, 48, 157–169. https://doi.org/10.1016/j.jmsy.2018.01.006

[31] Too, E. C., Li, Y., Njuki, S., & Liu, Y. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture, 161, 272–279. https://doi.org/10.1016/j.compag.2018.03.032

[32] Touvron, H. (2023). ResMLP: Feedforward networks for image classification with data-efficient training. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4), 5313–5324.

[33] Tseng, A. M., Shrikumar, A., & Kundaje, A. (2020). Fourier-transform-based attribution priors improve the interpretability and stability of deep learning models for genomics. bioRxiv. https://doi.org/10.1101/2020.06.11.147272

[34] Wang, Y., Li, H., & Zhao, X. (2023). Frequency-aware convolutional neural networks for image recognition. IEEE Access, 11, 30214–30227.

[35] Williams, M., & Li, R. (2018). Wavelet pooling for convolutional neural networks. Proceedings of the ICLR Workshop.

[36] Wolfert, J., Ge, L., Verdouw, C. N., & Bogaardt, M. J. (2017). Big data in smart farming: A review. Agricultural Systems, 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

[37] Yang, R., Zhu, S., Zheng, X., & Zeng, B. (2022). OL-DN: Online learning based dual-domain network for HEVC intra frame quality enhancement. arXiv. https://doi.org/10.48550/arXiv.2208.04661

[38] Yu, D., Wang, H., Chen, P., & Wei, Z. (2014). Mixed pooling for convolutional neural networks. In Lecture

Notes in Computer Science (Vol. 8834, pp. 364–375). Springer. https://doi.org/10.1007/978-3-319-11740-9_34

[39] Zargar, A. A. (2023). Tomato leaf disease dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6973251

[40] Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. arXiv. https://doi.org/10.48550/arXiv.1301.3557

[41] Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674