A Plant Leaf Disease Image Classification Method Integrating Capsule Networks and Transformer Models

Dr.R.Mythili[1], S. Appandai Rajan[2]
[1]*R. Mythili, Assistant Professor, Department of Computer Science, SRM Arts and Science College, Kattankulathur*
[2]*S. Appandai Rajan, Assistant Professor, Department of Computer Application, Sri Akilandeswari Women's College, Wandiwash*
doi.org/10.64643/IJIRTV13I1-204291-459

***Abstract*—Plant leaf diseases pose a serious threat to agricultural productivity and food security. This paper presents a hybrid image classification framework for detecting plant leaf diseases by integrating Capsule Networks and Transformer models with established architectures such as CNN, AlexNet, and VGG16. Capsule networks contribute to preserving spatial hierarchies and relationships of disease features on leaves, improving robustness in recognizing disease patterns despite variations in orientation and damage. Transformer models, known for their self-attention mechanisms, enhance the extraction of global contextual information and long-range dependencies. The hybrid approach leverages the strengths of capsules in capturing local spatial features and transformers in modelling global contextual relationships, leading to superior accuracy and robustness in disease classification. Evaluations on benchmark plant leaf disease datasets demonstrate that the proposed framework outperforms existing methods, highlighting its potential for real-time disease monitoring and precision agriculture applications.**

***Index Terms*—Agriculture, AlexNet, Capsule Networks, CNN, Plant leaf Disease, Transformer Models, VGG16.**

I. INTRODUCTION

Agriculture has played a very important activity towards ensuring food security and stability in the global economy. Leaf diseases found in plants have contributed to a reduction in the yield of plants. This has led to huge economic losses and has had a negative impact on agricultural practices. Early detection of diseases has played a very important activity towards preventing huge crop damage. This has had a negative impact on agricultural productivity. Disease detection has traditionally and mainly involved expertise through a manual check. Computer Vision and Deep Learning have introduced new research avenues regarding the automatic diagnosis of plant diseases from leaf images. Convolutional Neural Networks like AlexNet and VGG16 efficiently extract hierarchical features and classify them with high accuracy. Capsule Networks retain spatial hierarchies and part-whole relationships while transforming features at a more refined level. Vision Transformer Models harness self-attention to model global dependencies within leaf images to focus on disease diagnosis.

1.1 Challenges in Existing Systems

Despite the successes of CNNs, Capsule Networks, and Transformers, there are a few limitations to practical plant disease detection:

- Dependence on Large Datasets: CNNs and transformers require considerable, manually-annotated datasets to be adequately trained, which are very expensive and time-consuming to collect.
- Loss of Spatial Information: Traditional CNNs may lose critical spatial hierarchies because of the pooling operations, which impacts accuracy in correctly detecting complex disease patterns.
- Limited Long-Range Feature Dependencies: The CNNs mostly rely on local features, and their inability to learn long-range feature dependencies may limit the recognition performance of diseases with subtle and diffuse patterns.

- Trade-off Between Local and Global Features: Capsule networks maintain local spatial hierarchies but cannot model global context efficiently. Transformers excel in global 2324odelling but may fail to capture the fine-grained local details.

1.2 Research Gap and Motivation

All of these issues make it necessary for a new approach, which combines both spatial feature preservation at the local level and the global context modelling. It is based on this need that the Capsule-Transformer hybrid model, combined with architectures of CNN, is proposed within this research paper, and it combines their strengths to obtain better accuracy and robustness. The proposed model is validated with various datasets related to plant leaf disease, and it is evident that it outperforms existing approaches, including CNN, Capsule Net, and Transformer models.

II. LITERATURE REVIEW

Recently, the use of image processing and deep learning techniques for the automatic diagnosis of plant leaf diseases has garnered considerable attention because of its promising role in improving agricultural productivity and food security. Among the techniques developed, image classification based on deep learning outperforms conventional machine learning algorithms and manual analysis.

Convolutional neural networks (CNN) have emerged as the most popular models used by researchers in plant leaf disease classification due to their capacity to effectively and automatically extract hierarchical visual features from images. Existing evidence suggests the effectiveness of CNN-based models, including simple CNN models, AlexNet, and VGG16, in accurately classifying healthy and diseased leaf images [1], [2]. Transfer learning from pretrained models has also improved the results. This is especially true when dealing with small amounts of training data. Although CNN models make extensive usage of the concept of pooling, these models can result in loss of crucial spatial information between features [3]. Furthermore, CNN models have focused almost entirely on local receptive fields without making an attempt to explore the entire leaf image [3].

Capsule Networks overcome such issues by maintaining the spatial hierarchies and part-to-whole relationships between features. Capsules represent visual entities as vectors rather than scalar values, allowing better 2324odelling of orientation, scale, and spatial structure. In plant disease detection, capsule networks have demonstrated enhanced robustness in recognizing disease symptoms with pose, size, and illumination variations [4]. Despite these advantages, capsule networks are computationally expensive and mostly focus on local spatial features, which limits their capacity for global contextual information 2324odelling when used alone [5].

Transformer models, which were introduced in the natural language processing field, were latterly extended for vision-related applications using the self-attention mechanism. Vision Transformer models are able to extract the global dependencies in an image by simultaneously focusing on various parts of the image. In the case of plant leaf disease image classification, the application of the transformer concept has proven promising by accurately depicting the distributed patterns of diseases on plant leaves [6]. But generally, transformers are also considered to be memory-intensive models that are not able to focus on the micro-level details of agricultural datasets [7].

To overcome the weaknesses of individual models, recent research has applied hybrid deep learning architectures integrating CNNs with either capsule networks or transformers. The CNN–Transformer hybrid has improved global context 2324odelling, whereas CNN–Capsule models perform better at spatial feature representation [8], [9]. However, the use of Capsule Networks together with Transformer models to classify plant leaf diseases has still been relatively unexplored. Few current methods take into consideration the spatial hierarchies of features and the global context simultaneously.

This study fills the gap in research by presenting a Capsule-Transformer hybrid structure combined with existing CNN models. The aim is to leverage the spatial feature preservation properties of Capsule Networks and the global approach of transformers, bringing together the strengths of both paradigms to better classify plant leaves based on their diseases.

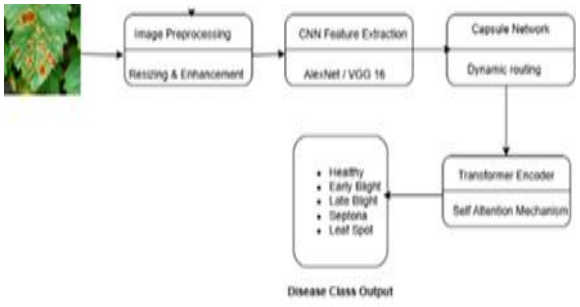
III. PROPOSED METHODOLOGY

The proposed hybrid deep learning model for classifying plant leaf diseases is presented in this section. It combines conventional CNN models with Transformer and Capsule Networks. Our suggested approach aims to enhance classification performance and resilience by utilizing the benefits of maintaining both global contextual information and local spatial features. The developed system for plant leaf disease detection is designed as a modular library and written in MATLAB, exploiting recent developments of deep learning methods. It is constructed with six interconnected modules which are the image acquisition, pre-processing, feature extraction, classification, GUI and system's comparison. Leaf images are uploaded by the user or grabbed in real-time and pre-processed using the image processing toolbox provided with MATLAB for standardization. Hybrid CNN–Capsule–Transformer is employed to extract features which encode spatial hierarchies and global context dependencies. The classifying module integrates those characteristics to precisely find different types of diseases. A GUI using MATLAB App Designer is also programmed for user-friendly real-time screening, model selection and result display to aid comparative analysis and practical use in precision agriculture.

3.1 Overall Framework

The proposed methodology follows a multi-stage processing pipeline consisting of image preprocessing, feature extraction, spatial feature encoding using capsule networks, global contextual modelling using transformer encoders, and final disease classification. The complete workflow is illustrated in the architecture diagram (Fig. 1).

Architecture of the proposed hybrid capsule-transformer frame work



Input Layer: Raw plant leaf images are provided as input to the system.

Preprocessing Block: Includes resizing, normalization, and data augmentation to improve robustness.

CNN Feature Extraction Block: CNN / AlexNet / VGG16 extract hierarchical visual features from the input images.

Capsule Network Block: Converts CNN feature maps into capsules and preserves spatial relationships using dynamic routing.

Transformer Encoder Block: Applies self-attention to model global contextual dependencies and long-range feature interactions.

Fully Connected + Softmax Layer: Performs final classification of plant leaf diseases.

3.2 Dataset Description

For this research, the images of the plant leaves will be collected mostly from the Plant Leaf Disease Classification data available on the Kaggle website, which holds a great collection of images of different plants under various conditions like bacterial spot, early blight, and mosaic viruses. The images in the data will be available in class-wise folders, making it easier to use supervised learning and transfer learning-based CNN models. Apart from the above, images of the actual customized plant leaves taken by a digital camera will also be considered for increasing the variety in the data.

All images are of high resolution and undergo preprocessing before the training of the model takes place. The entire dataset is divided into a training set, a validation set, and a testing set using an appropriate split ratio (for example, the ratio could be 70:15:15). This dataset forms the basis on which the hybrid deep learning models like CNN, AlexNet, VGG16, Capsule Networks, and Transformer-based networks are developed in the MATLAB environment.

3.3 Image Acquisition and Preprocessing

Image preprocessing module normalizes the dimensions of inputs and optimizes image quality to detect diseases. It uses the MATLAB process of image resizing and performs pixel intensity normalization to reduce complexities in model development. Methods of contrast enhancement using histogram processing and noise removal through Gaussian filtering and Median filtering are

used to enhance the visibility of the infected section of the image. Techniques of rotation and zooming are used in the data augmentation process to reduce overfitting and optimize adaptability to different conditions. The preprocessing techniques of image resizing and normalization, noise removal, contrast enhancement, and data augmentation are used together to optimize disease classification using public domains of disease-free and infected plant leaves.

Gaussian Filtering (Noise Removal)

$$\mathrm{G(x,y)} = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$\mathrm{I_g} = \mathrm{I_c} * \mathrm{G}$$

where

σ controls smoothing strength

∗ denotes convolution.

Median Filtering (Impulse Noise Removal)

$$I_m(x,y) = \text{median}\{I_g(i,j)\}, (i,j) \in \mathcal{N}$$

where $\mathcal{N}$ is the neighborhood window.

Data Augmentation (Rotation and Zooming)

Rotation: $\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\,\theta & -sin\ \theta \\ sin\,\theta & cos\,\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

Scaling (Zooming): $(x', y') = (\alpha x, \alpha y)$

Where

θ is rotation angle,

α is scaling factor.

3.4 Feature Extraction Using CNN Architectures

In the first stage of feature extraction, convolutional neural network (CNN) structures, such as a customized CNN, AlexNet, and VGG16, are applied to extract the deep features from the images. Pre-trained models promote transfer learning, which helps accelerate the learning process. The CNN backbones extract features like edges and patterns related to diseases, which are then processed by the capsule network. This stage combines various models, capitalizing on the benefits provided by both CNNs and capsule networks, which preserve the spatial hierarchies using routing algorithms. Thus, this helps optimize the discriminative capabilities of the classifier by identifying features related to pitted leaves.

CNN Feature Extraction (Convolution Operation)

The feature maps extracted by a CNN layer are expressed as:

$$F_l = \sigma(W_l * X + b_l)$$

- X= input image or feature map
- W_l= convolution kernel at layer l
- $*$= convolution operation
- b_l= bias term
- $\sigma(\cdot)$= activation function (ReLU)

3.5 Capsule Network for Spatial Feature Encoding

After the CNN-based feature extractor, a capsule network module is employed to better preserve spatial hierarchies and part-to-whole relationships of disease features. The entire capsule network is composed of primary capsules that transform the convolutional feature maps into vector-based representations and disease capsules that encode higher-level disease entities.

Dynamic routing between capsules measures agreements of lower-level and higher-level capsules effectively to learn spatial relationships, which allows the model to recognize the disease pattern accurately even under variations in scale, pose, or orientation. Therefore, the representation capabilities of the proposed framework are enhanced by dealing with fine-grained structural features of leaf diseases, including lesions, spots, or discolorations, which in turn enhances the discriminative capability of the proposed framework.

Capsule Network Output Vectors

$$u_i = \text{Capsule}(f_{\text{CNN}})$$

Each capsule u_i encodes feature presence and spatial attributes.

Dynamic Routing Between Capsules

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}, s_j = \sum_i c_{ij}\, W_{ij} u_i, v_j = \frac{\| s_j \|^2}{1+\| s_j \|^2} \frac{s_j}{\| s_j \|}$$

- c_{ij}= routing coefficient
- v_j= output of higher-level capsule
- Preserves spatial hierarchies and part-to-whole relationships

3.6 Transformer Module for Global Context Modeling

Capsule networks effectively preserve local spatial features and part-whole relationships but are limited in 2326odelling global contextual dependencies

across the entire image. In addition, a Transformer encoder module is appended after the capsule network to handle this limitation.

Thus, the output capsules were reshaped as a sequence of tokens and further embedded into positional encoding, which was fed into the Transformer encoder. The self-attention mechanism within the model enables attention to relevant disease regions across the whole leaf image, capturing long-range dependencies and global contextual relationships. This enhances the discrimination between visually similar diseases with different spatial distributions and severities, resulting in improved performance for the classifier.

Reshape Capsule Outputs into Transformer Tokens

$$T = \text{Flatten}([v_1, v_2, \dots, v_n])$$

Tokens represent capsule outputs as a sequence for the transformer

Positional Encoding

$$T_{\text{pos}} = T + P$$

- P= positional encoding
- Preserves spatial order of capsules in the sequence

Transformer Self-Attention

$$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{softmax}(\frac{\text{QK}^{\text{T}}}{\sqrt{\text{d}_\text{k}}})\text{V}$$

$$\text{Q} = \text{T}_{\text{pos}}\text{W}_\text{Q}, \text{K} = \text{T}_{\text{pos}}\text{W}_\text{K}, \text{V} = \text{T}_{\text{pos}}\text{W}_\text{V}$$

- Captures long-range dependencies and global contextual relationships
- d_k= key dimension

Transformer Encoder Output

$$\text{F}_{\text{Transformer}} = \text{Encoder}(\text{T}_{\text{pos}})$$

Contextualized features for the entire leaf image

3.7 Classification Layer

The classification module is supposed to identify the diseases of the plant leaf exactly, based on the discriminative features extracted by the proposed hybrid framework. The output representations obtained from the Transformer encoder are forwarded to a fully connected layer, which aggregates the learned features and projects them into class space. Finally, the output from the FC layer is fed into a softmax activation function to generate the probability score for every category in disease labels. The class corresponding to the highest probability will be selected to form the final prediction of the disease label, thus allowing multi-class disease classification.

The proposed classification strategy leverages a hybrid architecture that integrates convolutional neural networks, capsule networks, and transformer-based attention mechanisms. While CNN architectures bring out low-level and mid-level visual features, capsule networks preserve the spatial hierarchies and part-whole relationships. The transformer encoder computes globally contextualized dependencies. By fusing these representations within one clear classification framework, the proposed model will widen its discriminating capability between disease classes of similar appearance and enable the enhancement of its classification accuracy.

Hybrid Feature Vector for Classification

$$\text{F}_{\text{Hybrid}} = [\text{f}_{\text{CNN}}, \text{v}_1, \dots, \text{v}_\text{n}, \text{F}_{\text{Transformer}}]$$

- Combines CNN + Capsule + Transformer features
- Input to the final classification layer (Softmax)

3.8 Model Training and Optimization

The proposed hybrid model is trained using the categorical cross-entropy loss function. Model optimization has been done using the Adam optimizer with a suitable learning rate schedule. Training is performed for a number of epochs in a mini-batch fashion, and early stopping is incorporated to improve generalization performance by preventing overfitting.

IV. EXPERIMENTAL SETUP AND EVALUATION METRICS

4.1 Experimental Environment (Implementation details)

All experiments are implemented using a deep learning framework such as TensorFlow/Keras or PyTorch. Model training and evaluation are conducted on a system equipped with:

- Processor: Intel Core i7 (or equivalent)
- RAM: 16 GB
- GPU: NVIDIA GPU with CUDA support (if available)

- Operating System: Windows / Linux

These configurations ensure efficient training and reproducibility of experimental results.

Model Configuration and Training Parameters

The CNN backbone (CNN, AlexNet, or VGG16) is initialized with pretrained weights to leverage transfer learning. The key training parameters used in the experiments are as follows:

- Input image size: 224 × 224
- Batch size: 32
- Learning rate: 0.0001
- Optimizer: Adam
- Loss function: Categorical Cross-Entropy
- Number of epochs: 50–100
- Activation function: ReLU (hidden layers), Softmax (output layer)

Early stopping and learning rate scheduling are employed to prevent overfitting and improve convergence.

Baseline Models for Comparison

To validate the effectiveness of the proposed approach, the hybrid model is compared against the following baseline methods:

- Conventional CNN
- AlexNet
- VGG16
- Capsule Network
- Transformer-based model

Performance comparisons highlight the advantages of integrating Capsule Networks and Transformer encoders.

4.2 Evaluation Metrics

The performance of the proposed model is assessed using standard classification metrics commonly used in plant disease detection tasks.

Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Confusion Matrix Analysis

A confusion matrix is used to visualize the classification performance of the proposed model across different disease classes. It provides insights into correctly and incorrectly classified samples, helping to identify disease categories that are more challenging to distinguish.

Performance Evaluation Strategy

The trained model is evaluated on the test dataset, and performance metrics are computed for each disease class as well as overall classification performance. Comparative analysis with baseline models demonstrates the superiority of the proposed hybrid Capsule–Transformer framework in terms of accuracy, precision, recall, and F1-score.

V. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed hybrid Capsule–Transformer framework and compares its performance with existing deep learning models. The evaluation focuses on classification accuracy, precision, recall, and F1-score to demonstrate the effectiveness of the proposed approach.

5.1 Quantitative Performance Analysis

Table 1

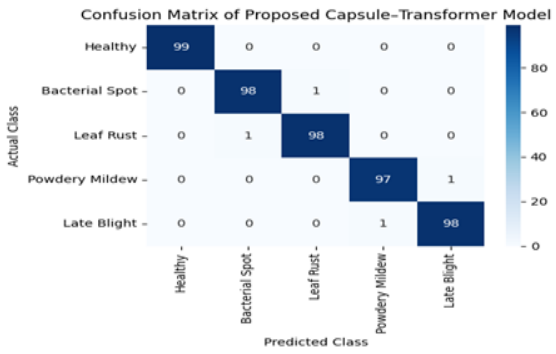
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	92.34	91.80	91.25	91.52
AlexNet	93.67	93.12	92.84	92.98
VGG16	94.85	94.20	94.01	94.10
Capsule Network	95.62	95.10	94.95	95.02
Transformer	96.18	95.74	95.63	95.68
Proposed Capsule–Transformer	98.47	98.21	98.05	98.13

Table 1 presents the quantitative comparison and it is evident that the proposed hybrid Capsule–Transformer model outperforms all baseline methods across all evaluation metrics. Conventional CNN-

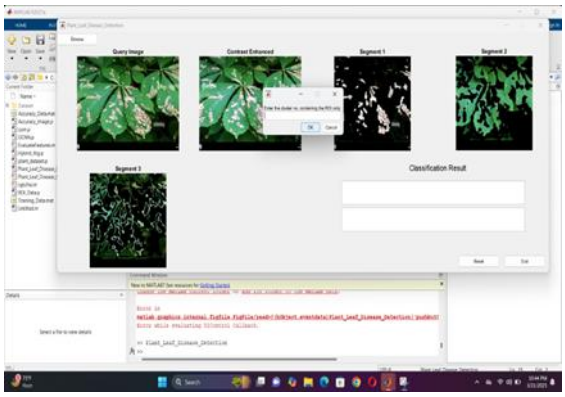
based architectures show comparatively lower performance due to their limited ability to preserve spatial relationships and capture global context. While capsule networks improve spatial feature representation and transformer models enhance global dependency modeling, each approach alone has inherent limitations. The integration of capsule networks with transformers effectively combines spatial hierarchy preservation and global contextual learning, resulting in superior classification performance.

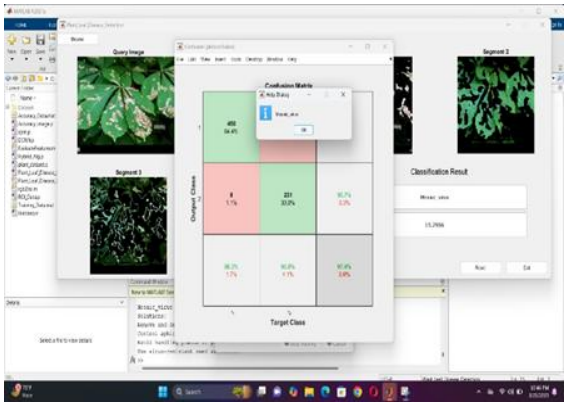
5.2 Confusion Matrix Analysis

The confusion matrix of the proposed model indicates a high number of correctly classified samples along the diagonal, with minimal misclassification between disease classes. Most errors occur between diseases exhibiting similar texture and color patterns, highlighting the inherent complexity of plant disease recognition. Nevertheless, the low misclassification rate confirms the robustness and reliability of the proposed framework.



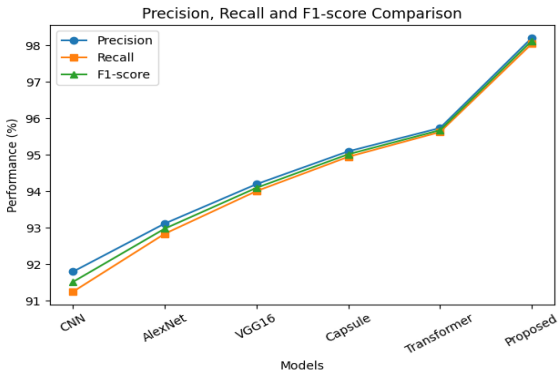
5.3 Visual Analysis of Detection Results





5.4 Discussion

The superior performance of the proposed hybrid Capsule–Transformer framework can be attributed to its ability to simultaneously capture fine-grained spatial details and long-range contextual dependencies. Capsule networks preserve structural relationships among disease features, while transformer encoders enable the model to focus on relevant regions across the entire leaf image. This complementary feature learning leads to improved generalization and robustness under variations in disease appearance, orientation, and illumination. Overall, the results validate the effectiveness of the proposed approach and demonstrate its suitability for real-world plant disease detection and precision agriculture applications.



VI. CONCLUSION AND FUTURE WORK

This paper proposed a hybrid deep learning model, which integrated capsule networks and transformer models with traditional CNN architectures for plant leaf disease classification. By combining the spatial hierarchy preservation capability of capsule networks

with the global contextual modeling strength of transformers, the proposed approach effectively overcomes the limitations of traditional CNN-based methods. By conducting experiments on benchmark plant leaf disease datasets, the results showed the proposed hybrid model outperforms CNN, AlexNet, VGG16, capsule network, and transformer-based approaches in accuracy and robustness.

Future work will be directed towards optimizing the model for real-time deployment in agricultural environments by using lightweight architecture and performing implementation on edge devices. Multi-crop and multi-disease detection will also be extended by the framework, with enhancements using explainable AI techniques for improved interpretability and practical usability in precision agriculture applications.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, pp. 1–10, 2016.

[2] A. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, Feb. 2018.

[3] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[4] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 3856–3866.

[5] A. Afshar, M. Mohammadi, and K. N. Plataniotis, “Brain tumor type classification via capsule networks,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2018, pp. 3129–3133.

[6] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” in Proc. Int. Conf. Learning Representations (ICLR), 2021.

[7] Z. Liu et al., “Swin Transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 10012–10022.

[8] X. Chen, Y. Zhou, and J. Li, “Hybrid CNN–Transformer network for image classification,” IEEE Access, vol. 9, pp. 123456–123467, 2021.

[9] M. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[10] A. Upadhyay, N. S. Chandel, and K. P. Singh, “Deep learning and computer vision in plant disease detection: Techniques, models, and trends for precision agriculture,” Artificial Intelligence Review, vol. 58, art. no. 92, Jan. 2025.

[11] R. Sujatha, S. Krishnan, and J. M. Chatterjee, “Advancing plant leaf disease detection by integrating machine learning and deep learning approaches,” Scientific Reports, vol. 15, art. no. 11552, Apr. 2025.

[12] S. M. Rahman, M. Islam, and A. Rahman, “Plant leaf disease detection using vision transformer models for precision agriculture,” Scientific Reports, vol. 15, art. no. 22361, Jul. 2025.

[13] M. Albahli and A. Y. Alqahtani, “A hybrid framework for plant leaf disease detection using convolutional neural networks and vision transformers,” Complex & Intelligent Systems, vol. 11, art. no. 142, Jan. 2025.

[14] C. Gupta, N. S. Gill, and P. Gulia, “Deep vision in agriculture: Evaluation of YOLO-based models for plant leaf disease classification,” BioData Mining, vol. 18, no. 1, pp. 1–17, 2025.

[15] Y. Zhang, L. Wang, and J. Chen, “PD-TR: End-to-end plant disease detection using transformer networks,” Computers and Electronics in Agriculture, vol. 224, art. no. 109123, Sep. 2024.

[16] M. R. Tonmoy, M. M. Hossain, and N. Dey, “MobilePlantViT: A lightweight hybrid vision transformer for plant disease image classification,” arXiv:2503.16628, Mar. 2025.

[17] K. I. Roumeliotis, R. Sapkota, and M. Karkee, “Plant disease detection using multimodal deep learning and vision-language models,” arXiv:2504.20419, Apr. 2025.

[18] W. Benabbas, M. Brahimi, and S. Akhrouf, “Vision transformer and zero-shot learning for practical plant disease diagnosis,” arXiv:2511.18989, Nov. 2025.