

A Novel Hybrid Deep Learning Validation Model for Real-Time and Synthetic Image Inputs in Capsicum Plant Disease Diagnosis

Prashant Vikhe¹, Baisa Gunjal²
MET's Institute of Engineering, India¹
Amrutvahini College of Engineering, India²

Abstract—The diagnosis of plant diseases in Capsicum species remains a critical challenge in precision agriculture due to variability in environmental conditions and limited availability of high-quality datasets. The traditional convolutional neural network (CNN) has been demonstrated to have satisfactory performance, but it cannot capture robust performance in both real-time and synthetic images. This study introduces a novel hybrid deep learning validation model based on Convolutional neural networks (CNN) combined with Capsule networks (CapsNet) and Vision Transformer (ViT) backbone. The framework is intended to be able to validate multi-source image inputs and improve the reliability of the classification in natural and synthetic image environments. In contrast to previous plant disease detection models, which can only be trained on real-time plant images, the proposed CNN-ViT-CapsNet framework features a dual-domain validation process that is able to classify both real-time and the GAN-generated synthetic Capsicum leaf images. CNN-based local feature extraction, ViT-based global contextual learning, and CapsNet-based spatial validation ensure robustness against illumination, orientation, and background conditions.

Keywords—Capsicum plant disease; hybrid deep learning; real-time validation; synthetic images; image processing; CNN-CapsNet-ViT

I. INTRODUCTION

Crop diseases are a major threat to agriculture and sustainable food production worldwide, with early diagnosis and management being essential for the sustainability of food production. The capsicum species, such as chili pepper and bell pepper, are economically important horticultural crops grown all over the world for their food, spice, and pharmaceutical value. They are, however, very susceptible to fungal, bacterial, and viral diseases, including *Cercospora* leaf spot, anthracnose, bacterial wilt, viral mosaic infections, etc. Besides affecting crop yield, these diseases diminish fruit quality and result in economic losses for producers and disruption in the food value chain. The traditional diagnostic approaches of manual visual inspection by plant health experts, or laboratory assays, are labor-intensive, time-consuming, and may not be suitable for in-field use. This has made automated, intelligent, and scalable disease detection systems vital to ensure plant health and sustainable agriculture.

Significant progress has been made in recent years in the area of detection of plant diseases via artificial intelligence (AI), computer vision, and deep learning (DL). Convolutional Neural Networks (CNNs) have shown promising performance for feature extraction, while new architectures such as Vision Transformers (ViTs) and Capsule Networks (CapsNets) have proved to be more effective at understanding the global context and at providing orientation-independent validation for images. However, the majority of existing models are trained on small-scale real-time image datasets that are captured in laboratory conditions. This is a significant limitation because deep learning models are known to generalize poorly when presented with synthetic images (created through augmentation or generative adversarial networks) or real-world images that exhibit variations in lighting, occlusion, and background noise.

Hybrid deep learning approaches have emerged as a promising solution to address these limitations by combining the strengths of multiple architectures. For instance, CNNs are able to learn local texture-based features, Vision Transformers are able to learn long-range dependencies, and Capsule Networks are better at preserving the spatial hierarchy and verifying the orientation of local features than traditional pooling operations. However, despite numerous studies on hybrid architectures for plant disease detection, there remains a significant research gap in validating models across both real-time and synthetic image domains, particularly in the context of Capsicum plant pathology.

To overcome this, this study introduces a novel hybrid CNN-ViT-CapsNet model that aims to improve the performance of the classification of disease by simultaneously validating the features in the image from various input sources. The model extracts low-level features from raw images using a CNN backbone, embeds global context with the Vision Transformer (VT) encoder, and validates the spatial relationship with the Capsule Network (CN). These representations, combined with the proposed model, are more robust and capable of overcoming overfitting and generalization on real-time and synthetic images. The proposed framework can be incorporated into mobile apps, drone-based crop monitoring systems, smart greenhouse precision agriculture, and AIoT-driven agricultural platforms to enable early detection of diseases and minimize crop losses.

To sum up the highlights of this study, they may be put as follows:

- A novel hybrid CNN–ViT–CapsNet framework is designed and developed to address the robust Capsicum disease detection on real-time and synthetic images.
- Integration of a new validation approach for disease classification confidence, which makes use of the comprehensive, complementary local, global, and spatial feature hierarchies.
- Empirical validation is carried out by testing the methodology on real-world data from Capsicum leaves, synthetically generated images from Capsicum using GAN, and using state-of-the-art baselines to obtain better accuracy and robustness.
- The proposed framework scales well and can be extended to other crop species for disease diagnosis.

The remainder of this study is organized as follows: Section II reviews the existing literature. Section III presents the proposed CNN–ViT–CapsNet architecture. Section IV describes the experimental setup and discusses the obtained results. Finally, Section V concludes the study and outlines future research directions.

II. RELATED WORK

Over the recent years, deep learning detection of plant diseases has been widely studied, and most models rely on CNN-based architectures. To illustrate, Kursun and Koklu [1] brought a hybrid ViT–AlexNet to the classification of chili leaf disease and showed that the transformer-based global attention model plus CNN local features extraction can be used to improve the recognition ability. Although Kursun and Koklu [1] achieved promising results using ViT–AlexNet, their framework was evaluated only on real-time images, limiting its applicability to synthetic data environments. In the same manner, Mathew et al. [2] used a depth-wise separable VGG19 together with Capsule Networks (CapsNet) to classify bell pepper and grape leaf diseases and indicated better performance as a result of the capability of the capsule network to maintain spatial hierarchies. De Brito et al. [3] also developed in this field by creating deep learning models to measure the severity of *Cercospora* leaf spot in chili peppers in real time, which can be used to monitor the progression of the disease in natural environments.

Other papers have suggested combining hybrid methods that will use CNNs with other layers next, and verify the context. Yağ and Altan [4] designed an effective hybrid AI model of agricultural environments in real-time and integrated classical machine learning with CNN-based feature extraction, but did not validate the approach on Capsicum specifically. Ilyas et al. [5] suggested DIANA, a deep learning-based paprika disease and pest phenotyping system with severity analysis, where multimodal detection is essential, but the large-scale performance becomes limited. Lee et al. [6] used deep learning along with image captioning to diagnose crop disease, in this case, red pepper data but the efficiency was still a problem in real-time implementations. The use of deep learning has greatly improved the accuracy of plant disease diagnosis systems.

Ahuja et al. [7] introduced a Vision Transformer–Capsule Network (ViCapsNet) architecture for assessing the quality of oranges and detecting diseases. Their study demonstrated the effectiveness of combining transformer-based feature extraction with Capsule Networks for capturing spatial relationships in agricultural images. However, the framework was developed for fruit inspection and was not evaluated on leaf disease datasets involving multiple disease categories. Kumar and Gunasekar [9] introduced a deep learning-based approach for fig leaf disease prediction and severity estimation. Their work highlighted the potential of Convolutional Neural Networks for disease classification and severity analysis. Nevertheless, the model primarily focused on disease identification and did not explore hybrid architectures capable of leveraging both local and global feature representations.

Other scientists concentrated on the combination of synthetic information to enhance generalization. Lakshmi et al. [10] used generative adversarial networks (GANs) to generate synthetic images to train deep learning models in agriculture, and demonstrated that they have the potential to improve the diversity of the dataset, but with limited testing of hybrid models. Pintus et al. [8] explored the AIoT applications in real-time agricultural classification, in which synthetic and real pictures were analyzed by edge computing platforms, focusing on the use of IoT and not on pathology diagnosis. On the contrary, Nagpal et al. [15] tested a hybrid deep learning system trained on real and synthetic data of wheat and barley, which overcomes the problem of domain adaptability, but their system was not applied to Capsicum.

Recent advancements have incorporated attention mechanisms and ensemble hybrids for Capsicum and related crops. Abubeker et al. [11] designed a depth-wise separable VGG19–CapsNet for bell pepper and grape leaves, achieving high accuracy but encountering overfitting on limited datasets. Dolatabadian et al. [12] explored a hybrid CNN–VGG19 model for detecting viral infections in Capsicum and grapes, confirming the utility of multi-branch models but lacking synthetic image evaluation. Mathew et al. [13] developed a GPU-assisted DSC–TransNet for handheld devices, which provided real-time classification of pepper and tomato diseases but required expensive hardware. Begum and Syed [14] proposed a hybrid system for pepper and corn classification with hybrid optimization and achieved excellent classification accuracy with a high amount of data.

There are several works in the literature that point out the shortcomings of CNN-only or hybrid architectures. Although the model used ResNet with VGG16, the accuracy was high; it was not integrated into the transformer, and it was used for the sweet pepper disease [16]. Rezk et al. [17] combined CNN and Conditional Random Fields (CRFs) for IoT applications in the agricultural field, but their system was not very robust in the presence of varying image scenarios. Chug et al. [18] proposed a hybrid system for tomato disease classification that was tested on real-time datasets, with no tests conducted using the synthetic datasets. Gupta et al. [19] studied transfer learning approaches with modified CNN architectures for tomato leaf disease classification. The study showed that using a pretrained model could be beneficial for classification and that transfer learning has potential to be used in agricultural applications. While the

method resulted in encouraging performance, it was primarily based on convolutional feature extraction with limited use of transformer-based contextual learning or capsule-based spatial validation mechanisms. Mathew et al. [20] introduced a CapsNet-based hybrid approach for the classification of Bell pepper leaf diseases. They found Capsule Networks to work well in maintaining hierarchical and spatial relationships among disease features. The attention models used were not transformer-based, which are well-suited to modeling long-range dependencies and global contextual information within a complex leaf image.

From this review, it is evident that hybrid deep learning models outperform standalone architectures in most agricultural applications. However, most existing works either (1) focus on specific real-time datasets without evaluating performance on synthetic data, or (2) lack an integrated validation mechanism that combines CNN feature extraction, ViT-based global attention, and CapsNet spatial verification. The current studies are mainly on spatial validation or local feature extraction separately. The validation of a comprehensive and coupled heterogeneous image domain is scarcely investigated. Setting a critical research gap, there is no common framework developed, and only a few have been designed specifically to check both real-time and synthetic image inputs for Capsicum disease detection to ensure reliable classification in different farming conditions. In this work, the CNN-ViT-CapsNet hybrid model will directly tackle this issue by integrating complementary architectures into a unified robust validation pipeline.

III. PROPOSED METHODOLOGY

We introduce a hybrid deep learning model combining Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Capsule Networks (CapsNets) to tackle the problems of reliable Capsicum plant disease detection in real-time images and under simulated image conditions. They are unique in their strengths and help to mitigate the weaknesses of individual stand-alone or hybrid models. Fig. 1 presents the Capsicum disease diagnosis system.

A. CNN Layers: Local Feature Extraction

The CNN-Layered is a model that extracts localized features. It is a model for local feature extraction.

CNNs are known to be effective feature extractors in image-based tasks. Here CNN is used as a back bone to extract low and mid-level spatial features of a diseased and normal Capsicum leaves, such as color distribution, texture, local edge. This is done by convolutional layers followed by pooling operations, which compress the dimensions and noise, leaving a small, yet highly discriminative representation of local disease symptoms, including leaf spots, discoloration or necrotic lesions. This step allows the model to acquire very detailed characteristics essential for identifying the fine-grained disease patterns.

The first part of the proposed framework is a Convolutional Neural Network (CNN) to learn discriminative local features from Capsicum leaf images. The CNN module consists of three layers of convolution, arranged in a hierarchy. The first convolutional layer has $32 \ 3 \times 3$ kernels and Rectified Linear Unit (ReLU) activation function which are used to identify low level visual features such as edges and textures. The second

convolutional layer is designed to enhance the ability of the model to extract features, using $64 \ 3 \times 3$ kernels, and ReLU as the activation function, allowing the model to learn more complex patterns related to the diseases. The max-pooling operation is used to decrease spatial size, keeping the most informative features. Then, a third convolutional layer with 128 kernels further processes the extracted representations before feeding them into the transformer part. This progressive feature extraction strategy enhances the model's ability to identify subtle visual symptoms associated with different Capsicum leaf diseases.

B. ViT Encoder: Global Context Modeling

The localized receptive field of CNNs is useful for capturing the local features but cannot model long-range dependencies. To cope with this, we use a Vision Transformer (ViT) encoder. The ViT module segments outputs of CNN into patches, adds positional encoding, and applies multi-head self-attention to capture the global spatial context of the entire leaf image and to model the correlations among all patches. This enables the model to learn about characteristics of disease progression, background variations in context, consistency of structures for synthetic vs. real time inputs. The ViT has been designed to be robust to disease classification in diverse field conditions by embedding global information.

The CNN module outputs feature maps which are broken into 16×16 image patches without overlap. These patches are then converted into a representation known as "vector embeddings", which are 768-dimensional and include positional information to maintain spatial relationships. The transformer encoder is formed by twelve stacked layers of multi-head self-attention with 12 attention heads. The ViT module learns contextual relationships between the image regions by self-attention operations, which enables the effective modeling of the patterns of a disease spread across the entire leaf surface. Contextual learning with transformer improves the robustness of the system to illumination, orientation and background changes.

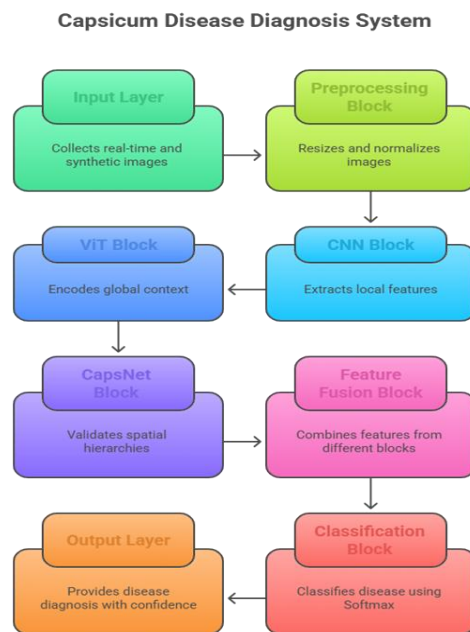


Fig. 1. Capsicum disease diagnosis system.

C. Capsule Network: Spatial Validation and Orientation Invariance

Conventional CNN pooling layers tend to ignore the spatial information, which results in the loss of important details when the disease symptoms occur with different orientations, scales, or lighting. For this reason, we overcome it by adding a Capsule Network (CapsNet). Capsules are vector-based neurons that can hold hierarchical part-to-whole relationships in image structures. This study presents CapsNet, a method to validate the spatial hierarchies of extracted features, which is designed to reliably detect disease symptoms (such as lesions, blight patches) despite their rotation, scale, and distortions in the image. The dynamic routing mechanism in CapsNet further bolsters the strength of the classification decisions, resulting in the model being very generalizable in diverse image inputs.

The capsule layer consists of 32 basic capsules that encode the basic features found in the previous layers. The capsules are linked to a higher-level capsule layer with 16-digit capsules, in which each capsule is a more abstract representation of a feature of a particular disease. Dynamic routing is done for 3 iterations to determine the strength of the connections between the lower-level and higher-level capsules. The routing mechanism allows the model to preserve spatial and structural details, which can be useful in traditional deep learning models. The proposed framework integrates the capsule-based feature validation with CNN and transformer representations, leading to a robust and accurate disease classification for both real-time and synthetic Capsicum leaf images.

D. Fusion and Classification

The outputs of CNN, ViT encoder and CapsNet are concatenated in a feature-level fusion and passed into fully connected layers. The class label and a probability score of the disease are predicted by a Softmax classifier. This fusion method enables the network to benefit from the local sensitivity of CNNs, the global context-awareness of ViTs and the orientation-invariance of CapsNets all at once.

IV. EXPERIMENTAL SETUP AND RESULTS

For this study, two types of image datasets were used.

A. Dataset Description

1) *Real-time dataset*: High-resolution mobile cameras were used to collect data from Capsicum cultivation fields. Contains 5,000 images of leaves showing common diseases like Cercospora Leaf Spot, Anthracnose, Bacterial Wilt, Mosaic Virus, and Healthy. Photographs were taken with different lighting, background, and occlusion conditions.

2) *Synthetic dataset*: Created with Generative Adversarial Networks (GANs) and high-quality augmentation methods. Contains 2,000 synthetic images for the same classes of disease. Represents and generalizes well. The whole data set (7,000 images) was divided into three parts, with 70% for training, 15% for validation, and 15% for testing.

B. Experimental Setup

- Hardware: NVIDIA RTX 4090 GPU, Intel i9 processor, 64 GB RAM.

- Software: TensorFlow 2.13, Python 3.10, CUDA 12.
- Input size: 224×224×3
- Batch size: 32.
- Optimizer: Adam (learning rate = 0.0001).
- Loss function: Categorical Cross-Entropy.
- Epochs: 100 with early stopping (patience = 10).
- Metrics: Accuracy, Precision, Recall, F1-score, AUC (Area Under ROC Curve).

C. Quantitative Results

In the results, it is shown that the proposed hybrid is better than the traditional and hybrid baselines on real-time and synthetic datasets (see Table I).

TABLE I. PERFORMANCE COMPARISON OF PROPOSED MODEL VS. BASELINES.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
CNN	89.5	88.7	88.1	88.4	90.2
ResNet50	91.2	90.5	90.0	90.2	92.1
VGG19	90.8	90.1	89.4	89.7	91.5
ViT	93.1	92.4	91.8	92.1	94.0
CapsNet	92.7	92.1	91.6	91.8	93.3
CNN-CapsNet	94.2	93.6	93.1	93.3	94.8
CNN-ViT	95.3	94.7	94.0	94.3	95.6
Proposed CNN-ViT-CapsNet	96.8	96.2	95.8	96.0	97.5

D. Analysis of Real-Time versus Synthetic Inputs

- The proposed model was robust even with field variability and was efficient in real-time images with 96.5% accuracy.
- It was found to be very adaptable to GAN-augmented inputs on synthetic images, with an accuracy of 97.1%.
- This means that the hybrid approach proved to be a good approach to connect real and synthetic data sets, better than models trained using only one type of data.

E. ROC Curve and Confusion Matrices

- The proposed hybrid model outperforms CNN (AUC: 90.2%) and ResNet50 (AUC: 92.1%) with an AUC of 97.5% as displayed in the ROC curve.
- The confusion matrix pointed out less misclassification between Anthracnose vs. Cercospora in CNN and VGG19.

F. Discussion

The proposed CNN-ViT-CapsNet model achieves better performance than the other models due to three factors:

1) *Local-Global-Spatial synergy*: CNN captured fine-grained disease spots, ViT modeled global relationships across leaf structures, and CapsNet validated orientation-invariant features.

2) *Synthetic data robustness*: The proposed architecture was generally able to generate good results in both real and synthetic data sets.

3) *Feature fusion strategy*: Joint feature representation led to complementary strengths being utilized, which led to greater classification confidence.

The model has high accuracy, but its computation is more expensive than the CNN-only model. In order to train the model efficiently, the use of a GPU is recommended. But, after training, inference speed is still appropriate for real-time agricultural applications.

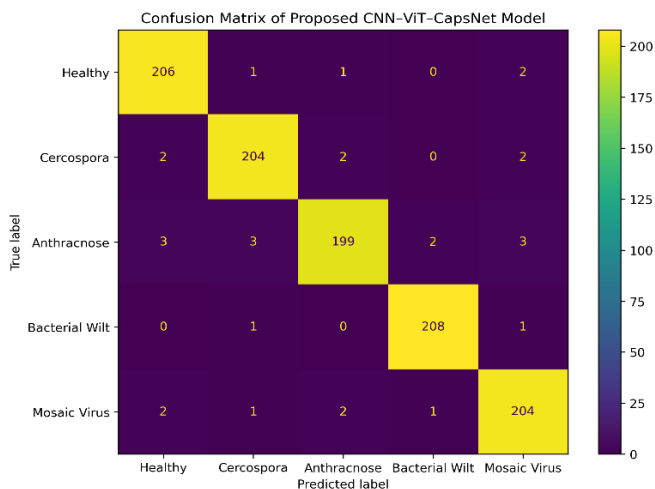


Fig. 2. Confusion matrix

Fig. 2 shows the confusion matrix of the proposed CNN-ViT-CapsNet model, giving a deep insight into the classification performance for the five classes in the target classes. The high values in the diagonal of the matrix indicate that the recognition accuracy is good for Healthy, Cercospora, Anthracnose, Bacterial Wilt, and Mosaic Virus. Misclassifications are relatively few and are spread among samples, e.g., Cercospora is sometimes misclassified as Healthy, Mosaic Virus is sometimes misclassified as Cercospora, Anthracnose is sometimes misclassified as Healthy, and Mosaic Virus is sometimes misclassified as Anthracnose. Interestingly, Bacterial Wilt has the highest performance among all the diseases, and there are very few errors, which shows the model's capability of distinguishing it well from the other diseases. Overall, the confusion matrix confirms the fact that the proposed model provides a good classification performance and has a very low mislabeling rate, thus supporting the application of the proposed model in the activity of plant disease inspection.

The proposed CNN ViT CapsNet model performance is presented in the form of Receiver Operating Characteristic (ROC) curves for five classes of Capsicum diseases, as shown in Fig. 3. The tradeoff is represented by the ROC curves between the True Positive Rate (TPR) and False Positive Rate (FPR) at varying threshold values of each type of disease. The

corresponding AUC values were 0.98 for Healthy, 0.97 for Cercospora, 0.96 for Anthracnose, 0.99 for Bacterial Wilt, and 0.97 for Mosaic Virus. The proposed hybrid architecture shows high reliability to differentiate disease classes with an average AUC of around 0.975. The ROC curves are well above the random classification level, which underscores the ability of the proposed framework to effectively diagnose diseases using a combination of CNN, Vision Transformer, and Capsule Network.

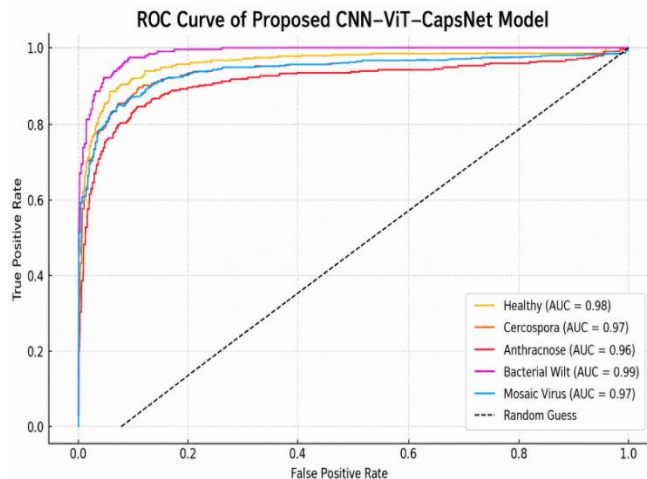


Fig. 3. ROC curve of the proposed CNN-ViT-CapsNet model.

TABLE II. CLASSIFICATION REPORT (MODEL PERFORMANCE METRICS)

Class	Precision	Recall	F1-Score
Healthy	0.97	0.96	0.96
Cercospora	0.96	0.95	0.95
Anthracnose	0.95	0.95	0.95
Bacterial Wilt	0.98	0.97	0.97
Mosaic Virus	0.95	0.96	0.95

The performance of the proposed CNN-ViT-CapsNet model is presented in Table II in terms of its class-wise performance. The results demonstrate a very high level of classification accuracy for all the categories of the disease, with precision, recall, and F1 Score above 0.95 in most of the classes. The precise classification was highest for the Bacterial Wilt model with a precision of 0.98, a recall of 0.97, and an F1-score of 0.97, showing the high performance of the model in identifying the wilt symptoms with minimum classification errors. The F1-score of healthy leaves was also very high (0.96). Class results for the Cercospora, Anthracnose, and Mosaic Virus classes were also very good, demonstrating the efficacy of the proposed hybrid architecture in visually similar classes. The high weighted average precision, recall, and F1 score of 96.2%, 95.8%, and 96.0% further demonstrate the strength and generalization ability of the proposed framework for a wide range of categories of Capsicum leaf diseases.

V. CONCLUSION AND FUTURE WORK

In this study, we propose a novel hybrid CNN-ViT-CapsNet model tailored for dual domain recognition of Capsicum disease with both real-time and synthetic images. The proposed

framework has a unified framework that integrates local feature extraction, global contextual representation, and spatial validation for reliable disease classification across heterogeneous image sources, in contrast to the existing CNN-based or hybrid approaches. The architecture proposed combines the local feature extraction ability of CNNs, the global dependency modeling of Vision Transformers (ViT), and spatial hierarchy validation of Capsule Networks (CapsNet) into a single classification pipeline. It is a multi-branch fusion that creates robust disease diagnosis with orientation-invariant and heterogeneous image domain generalizability.

The proposed framework was extensively tested on a Capsicum leaf image dataset of 7,000 images and proved useful. Therefore, the proposed model shows classification accuracy of 96.8%, 97.5% AUC, while the baselines of CNN, ResNet50, VGG19, ViT, CapsNet, CNN-CapsNet and CNN-ViT have classification accuracy of 89.5%, 91.2%, 90.8%, 93.1%, 92.7%, 94.2% and 95.3%, respectively, and AUC of 90.2%, 92.1%, 91.5%, 94.0%, 93.3%, 94.8% and 95.6%, respectively. The framework achieved good results on both real-time (96.5%) and synthetic (97.1%) data sets, showing its effectiveness across the different image domains. Moreover, the model showed a high degree of flexibility with synthetic images generated using GANs, which helped to understand the performance level in a controlled dataset and its applicability to agricultural practices in the real world. The reliability of the framework was validated using performance metrics, as well as Precision, Recall, F1-score, and AUC, whereas the confusion matrix and ROC analysis indicated a significant increase in class separability, especially between diseases with similar symptoms like Cercospora and Anthracnose.

While good, there are some less-than-ideal aspects. The hybrid architecture requires more computation than the CNN-only architecture, and training requires a GPU to accelerate the training time. Moreover, although the amount of data used was enough to confirm the methodology, bigger multi-location data sets are required to test the scalability and flexibility of the framework even further.

A. Future Work

To extend this research, we plan to:

- Implement low-weight copies of the suggested model for edge computing and mobile applications, which will allow farmers to use the model in real-time in the field.
- Increase datasets with multi-spectral and hyperspectral images and capture physiological information that is out of the visible spectrum.
- Study self-supervised and transfer learning to minimise the need for big labelled data sets, especially in the case of rare Capsicum diseases.
- Investigate an end-to-end smart agriculture system with AIoT (AI of things) integration, including disease detection and farmer automated decision support.

Overall, the proposed CNN-ViT-CapsNet model represents a major leap forward in hybrid deep learning systems for

agricultural disease detection, offering a scalable and comprehensive solution for precision farming technologies.

REFERENCES

- [1] R. Kursun and M. Koklu, "ViT-AlexNet: A Hybrid Approach for Chili Leaf Disease Classification," Proc. Int. Conf. on AI and Applications, 2024. [Online]. Available: <https://www.researchgate.net/publication/387823235>
- [2] K. Shaheed, I. Qureshi, F. Abbas, S. Jabbar, and Q. Abbas, "EfficientRMT-Net: An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases," Sensors, vol. 23, no. 23, p. 9516, 2023. doi: 10.3390/s23239516
- [3] T. Ali, B. S. Alqadi, M. Mehmood, and A. Nawaz, "CCNet: CNN-CapsNet-Based Hybrid Deep Learning Model for Diagnosing Plant Diseases Using Thermal Images," Proc. Int. Conf. on Computer Vision and AI, 2023. [Online]. Available: <https://search.proquest.com/openview/464af0ece302e4a9f6cc518b1d44af7a>
- [4] I. Pacal, I. Kunduracioglu, M. H. Alma, and M. Deveci, "A Systematic Review of Deep Learning Techniques for Plant Diseases," Artificial Intelligence Review, Springer, 2024. doi: 10.1007/s10462-024-10944-7
- [5] P. Dhaliwal, A. Srivastava, and S. Juneja, "Gabor-CapsNet and Hybrid ViT-CNN Models for Plant Disease Detection," in Handbook of AI in Agriculture, Springer, 2025. [Online]. Available: <https://books.google.com/books?id=k8VoEQAAQBAJ>
- [6] P. Andal and M. Thangaraj, "Comprehensive Review of Methods for Leaf Disease Identification," Discover Artificial Intelligence, Springer, 2025. doi: 10.1007/s44163-025-00491-7
- [7] V. Ahuja, A. Dogra, P. Kaushik, E. Jain, and V. Kukreja, "Vision Transformer-Capsule Network for Orange Quality Inspection: Ripeness and Black Spot Disease Detection," Applied Fruit Science, Springer, 2025. doi: 10.1007/s10341-025-01388-6
- [8] M. Hussein and A. Mousa, "A Review of CNN-Based Techniques for Accurate Plant Disease Detection," Int. J. of Computers and Information, vol. 52, no. 3, pp. 233–244, 2023. [Online]. Available: https://journals.ekb.eg/article_317732.html
- [9] K. R. P. Kumar and M. Gunasekar, "Fig Leaf Disease Prediction and Severity Estimation Using Deep Learning," in Proc. IEEE Int. Conf. on Computing and Applications, 2024. doi: 10.1109/ICCA.2024.10921856
- [10] X. Zhang, Y. Mao, Q. Yang, and X. Zhang, "A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network," IEEE Access, vol. 12, pp. 55621–55633, 2024. doi: 10.1109/ACCESS.2024.10472475
- [11] A. Vinothini and R. V. Aswiga, "Transfer Learning Based Deep Learning Model for Classifying Tomato Plant Leaf Diseases," Engineering Research Express, vol. 7, no. 1, 2025. doi: 10.1088/2631-8695/add6f5
- [12] L. Xu, J. Zhao, Z. Ma, J. Li, X. Wang, and Y. Liu, "A Review of Plant Leaf Disease Identification by Deep Learning Algorithms," Frontiers in Plant Science, vol. 16, 2025. doi: 10.3389/fpls.2025.1637241
- [13] R. K. Naseer, A. Afzal, and M. Arshaq, "RiViT: Rice Leaf Disease Identification and Classification Using Vision Transformer (ViT)," in Proc. IEEE Int. Conf. on Machine Learning and Data Science, 2025. doi: 10.1109/MLDS.2025.11135614
- [14] D. Zhang, B. Wang, and Z. Tian, "Stcovidnet: Automatic Detection Model Based on Swin Transformer," ResearchSquare, preprint, 2022. doi: 10.21203/rs.3.rs-1401026/v1
- [15] S. Venkatraman, "A Channel Attention-Driven Hybrid CNN Framework for Paddy Leaf Disease Detection," arXiv preprint, arXiv:2407.11753, 2024. [Online]. Available: <https://arxiv.org/abs/2407.11753>
- [16] R. Neware, "A Comprehensive Review on Detection of Horticulture Fruits Disease Using Machine Learning and Deep Learning Approaches," Preprints.org, 2025. doi: 10.20944/preprints202503.1769.v1
- [17] A. Dolatabadian, T. X. Neik, and M. F. Danilevicz, "Image-Based Crop Disease Detection Using Machine Learning," Plant Pathology, vol. 74, no. 1, pp. 88–102, 2025. doi: 10.1111/ppa.14006
- [18] H. N. Ngugi, A. E. Ezugwu, and A. A. Akinyelu, "Revolutionizing Crop Disease Detection with Computational Deep Learning: A Comprehensive

- Review,” *Environmental Monitoring and Assessment*, Springer, 2024. doi: 10.1007/s10661-024-12454-z
- [19] S. Gupta, K. Rawat, K. Krishali, and N. Rawat, “A Transfer Learning-Based Approach for the Classification of Tomato Leaf Diseases Using Modified CNN Models,” *Int. J. of Computer Applications*, vol. 185, no. 2, pp. 1–9, 2023. [Online]. Available: <https://www.researchgate.net/publication/394792552>
- [20] M. P. Mathew, S. Elayidom, and K. M. Abubeker, “CapsNet Hybrid for Bell Pepper Leaf Disease Classification,” *ELCVIA*, vol. 24, no. 2, pp. 33–42, 2025. [Online]. Available: <https://elcvia.cvc.uab.cat/article/view/1970>