Color-Aware Tomato Leaf Disease Classification

Hanzalah Sohail Faraz **Hussain Sulaiman Zia**

https://github.com/Hussain-SZ/DL-Project---Group-6

Abstract

Accurate plant disease classification is essential for ensuring food security and improving crop yields, especially in agriculture-dependent regions like Pakistan. In this study, we propose a deep learning pipeline for classifying tomato leaf diseases using RGB and Lab color space fusion. Starting from a dual-branch InceptionV3 baseline, we progressively enhance performance through the integration of EfficientNet backbones, extensive data augmentation, and attention mechanisms using CBAM. Evaluated on the PlantDoc dataset, our best model achieves a test accuracy of 55.63%, significantly outperforming our baseline (33.11%). We further present class-wise performance metrics and analyze failure cases to identify key limitations such as background clutter and class imbalance. Our work lays the groundwork for future extensions involving segmentation and natural language explanations to improve accessibility for farmers.

1. Introduction

Agriculture remains the backbone of Pakistan's economy, with a significant portion of the population relying on crop production for their livelihood. Among various crops, tomatoes hold substantial economic importance. However, tomato crops are susceptible to a range of diseases that can drastically reduce yield and quality. Timely and accurate identification of these diseases is crucial to mitigate losses and ensure food security.

Recent advancements in deep learning have shown promise in automating plant disease detection. Convolutional Neural Networks (CNNs) have been extensively utilized for this purpose, offering automated feature extraction and classification capabilities. For instance, Dey et al. (2024) demonstrated the efficacy of pre-trained CNN architectures like AlexNet, VGG16, and VGG19, achieving impressive testing accuracies exceeding 95% on the PlantVillage dataset (Dey et al., 2024). Similarly, Harakannanavar et al. (2022) employed machine learning techniques, incorporating image preprocessing methods such as Histogram Equalization and K-means clustering for segmentation, to detect tomato leaf disorders (Harakannanavar et al., 2022).

While these studies highlight the potential of deep learning in plant disease detection, they predominantly rely on the PlantVillage dataset. This dataset comprises images captured under controlled conditions, featuring individual leaves against uniform backgrounds. Such idealized images lack the variability and complexity present in real-world scenarios, limiting the generalizability of models trained solely on this data. In contrast, the PlantDoc dataset offers images collected in natural settings, encompassing diverse backgrounds, lighting conditions, and multiple leaves per image. This makes PlantDoc more representative of the challenges faced by farmers in real-life situations.

Building upon the insights from previous studies, our project aims to develop a robust model for tomato leaf disease classification using the PlantDoc dataset. Our baseline model employs a **TwoBranchInceptionV3** architecture, inspired by the work of Schuler et al. (Schuler et al., 2022), which we iteratively enhance through various improvements. These include integrating **EfficientNet** pretrained on ImageNet1K_V1 (Tan & Le, 2019), incorporating RGB+LAB color spaces, applying advanced image augmentation techniques, and embedding **Convolutional Block Attention Modules (CBAM)** to focus on salient features (Woo et al., 2018).

In the subsequent sections, we detail our methodology, present the results of our experiments, discuss the implications of our findings, and outline potential directions for future research.

2. Methodology

2.1. Baseline: Two-Branch InceptionV3 Model

Our baseline architecture is inspired by the ColorAware Two-Branch Inception model. The core idea is to exploit the perceptual benefits of the CIELab color space by separating luminance and chrominance information into two processing branches.

We begin by converting each input image from RGB to

Lab color space. The L (luminance) channel is separated from the AB (chrominance) channels. Each is normalized independently and then processed through a dedicated convolutional branch. The L and AB branches each consist of a convolutional layer followed by batch normalization and ReLU activation. The resulting feature maps are concatenated to form a combined representation, which is then passed to an InceptionV3 model.

We modify the input layer of InceptionV3 to accommodate the increased channel dimensionality. The model concludes with a fully connected classification head that maps to the number of disease classes. This two-branch architecture enhances the model's ability to extract meaningful features from both brightness and color cues, which is critical for fine-grained classification of plant diseases.

2.2. Improvement 1A: RGB-Lab Fusion with EfficientNet

While the Lab color space provides useful perceptual features, we observed that background artifacts were not well captured. To address this, we proposed a fused input representation by concatenating the original RGB channels with the Lab channels, yielding a six-channel input.

To effectively process this fused representation, we adopted EfficientNet-B0 as the backbone due to its efficiency and strong performance on image classification tasks. Initially, the model was loaded with pretrained weights from ImageNet to leverage the knowledge learned from large-scale image datasets. The first convolutional layer of EfficientNet was modified to accept six input channels instead of the standard three, while the rest of the architecture remained unchanged. Additionally, the classification head was reinitialized to match the number of target classes for our specific task.

After this modification, we fine-tuned the model on our specific dataset to adapt the pretrained features to our problem domain. This RGB-Lab fusion allowed the model to capture complementary color cues and improve generalization on diverse background conditions.

2.3. Improvement 1B: Data Augmentation

Given the limited size of the PlantDoc dataset, we introduced a strong data augmentation pipeline to enhance generalization. The RGB input pipeline included random resized cropping, horizontal and vertical flips, color jittering, affine transformations, perspective distortion, and probabilistic Gaussian blur. The Lab inputs were resized to maintain spatial alignment but otherwise remained unaugmented.

This augmentation strategy significantly increased data diversity and improved model robustness to lighting, pose, and background variations.

2.4. Improvement 2: Attention via CBAM and Residual Learning

To further refine feature learning, we integrated attention mechanisms using Convolutional Block Attention Modules (CBAM). These modules apply both channel and spatial attention to the convolutional features, allowing the network to emphasize the most informative parts of the image.

We inserted the first CBAM block after the final convolutional features of EfficientNet. This was followed by a custom residual block consisting of two 3×3 convolutional layers with batch normalization and ReLU activation. The output of this residual block was added back to the input to preserve gradient flow and enhance deep feature learning. A second CBAM block was then applied to the residual-enhanced features.

The final feature map was passed through the average pooling layer and a modified classification head. As before, the first convolutional layer of EfficientNet was adjusted to accept the six-channel RGB-Lab input.

This combination of attention and residual learning allowed the model to focus on disease-specific regions of the leaf while maintaining rich hierarchical features. The attention modules improved interpretability and feature localization, which is crucial in real-world plant disease diagnosis.

3. Results

Table 1 summarizes the classification performance of our four experimental models on the Tomato subset of the PlantDoc dataset. As expected, our baseline model—TwoBranchInceptionV3—achieved the lowest test accuracy of 33.11%. While it performed well on the cleaner PlantVillage dataset (99%), its generalization to real-world images suffered due to background clutter, lighting variation, and limited diversity in training samples.

To address these issues, our first improvement (EfficientNet + RGB+Lab without augmentation) replaced the InceptionV3 backbone with EfficientNet-B0 and fused RGB and Lab representations into a six-channel input. This modification alone increased accuracy to 47.68%, highlighting the benefit of richer color representation and pretrained feature extractors.

In the next stage, we introduced extensive data augmentation techniques including flipping, cropping, jittering, and affine transformations. These transformations significantly expanded the diversity of our dataset and further boosted performance to 53.64%. The model became notably more resilient to background noise and color distortions.

Finally, we incorporated CBAM attention modules into the EfficientNet backbone. This helped the model focus more effectively on disease-relevant regions in both color and spatial dimensions. The final accuracy reached 55.63%, representing a 22-point gain over the baseline. This demonstrates that channel and spatial attention mechanisms are effective even in relatively small and noisy datasets.

Table 1. Model Comparison on PlantDoc (Tomato Class)

MODEL	AUGMENTATION	ACCURACY (%)
TWOBRANCHINCEPTIONV3	NO	33.11
EFFICIENTNET + RGB+LAB	NO	47.68
EFFICIENTNET + RGB+LAB	YES	53.64
EFFNET+RGB+LAB+CBAM	YES	**55.63**

Table 2. Class-wise Precision, Recall, and F1-Score of Best Model

CLASS	PRECISION	RECALL	F1-SCORE
TOMATO LEAF YELLOW VIRUS	0.48	0.87	0.62
TOMATO LEAF MOSAIC VIRUS	0.30	0.27	0.29
TOMATO LEAF	0.59	1.00	0.74
TOMATO MOLD LEAF	0.86	0.32	0.46
TOMATO LEAF BACTERIAL SPOT	0.41	0.41	0.41
TOMATO SEPTORIA LEAF SPOT	0.50	0.53	0.52
TOMATO EARLY BLIGHT LEAF	0.62	0.28	0.38
TOMATO LEAF LATE BLIGHT	0.83	0.83	0.83
OVERALL ACCURACY		**55.63%**	
MACRO AVERAGE	0.57	0.56	0.53
WEIGHTED AVERAGE	0.59	0.56	0.54

Table 2 provides a detailed breakdown of precision, recall, and F1-score for each tomato disease class in the PlantDoc test set. The model performs exceptionally well on classes such as **Tomato leaf late blight** (F1 = 0.83) and **Tomato leaf** (F1 = 0.74), which have strong visual characteristics and higher inter-class separation.

However, the model struggles with visually similar or underrepresented classes. For example, **Tomato leaf mosaic virus** exhibits both low precision (0.30) and recall (0.27), due to limited training examples and confusing texture patterns. Similarly, **Tomato mold leaf** has high precision (0.86) but low recall (0.32), indicating the model is confident when it does predict the class, but fails to identify it consistently.

The macro average F1-score of 0.53 and weighted average F1-score of 0.54 highlight the dataset imbalance and the challenge of generalizing across multiple diseases. These results support our decision to explore segmentation, background suppression, and potentially class balancing in future work.

4. Discussion

Our experiments show that combining RGB and Lab color spaces, applying data augmentation, and introducing attention mechanisms such as CBAM substantially improve classification performance on real-world leaf images. The progression from 33.11% (baseline) to 55.63% (final model) demonstrates the effectiveness of these modifications, especially in handling complex color variations and background noise.

However, several limitations remain. We observed that standard architectures like ResNet18 (He et al., 2015) and MobileNetV2 (Howard et al., 2017) consistently underperformed on the PlantDoc dataset, with accuracies below 40%. These models, while efficient, may lack the representational capacity needed to handle fine-grained disease patterns in real-world conditions, especially when leaf symptoms are subtle or the background is cluttered. This reinforces the need for richer feature extraction pipelines and tailored preprocessing techniques.

Another significant limitation is the lack of spatial awareness. Our current pipeline classifies full images without explicitly focusing on the diseased regions. Misclassifications often occurred in cases where leaves were partially occluded, poorly lit, or surrounded by non-plant objects such as soil or plastic. Figure 1 will highlight such failure cases. This suggests that integrating a segmentation module or a leaf detector could help isolate the regions of interest and improve classification accuracy.

Additionally, while augmentations expanded the effective dataset size and improved generalization, they could not fully compensate for the limited number of labeled images per class. This was especially evident in low-performing classes like Tomato leaf mosaic virus, which had both low recall and few training samples.

Further, we plan to incorporate a segmentation pipeline using models like U-Net to extract clean leaf masks before classification.

Finally, to enhance real-world utility, we aim to integrate a Large Language Model (LLM) that explains predictions in farmer-friendly language. This would allow the model to output not just disease labels, but also treatment suggestions and urgency levels in local languages, making the system more actionable for agricultural stakeholders.

5. Conclusion

In this work, we proposed a deep learning-based pipeline for plant disease classification focused on tomato leaves using a dual color space approach (RGB + Lab). Starting from a Two-Branch InceptionV3 baseline, we progressively improved accuracy through EfficientNet backbones, advanced augmentations, and CBAM attention modules. Our final model achieved 55.63% accuracy on the PlantDoc test set—over 22 percentage points higher than our baseline—demonstrating the effectiveness of color-space fusion and attention in noisy, real-world agricultural settings.

Figure 1. Sample predictions from our CBAM-based EfficientNet model.

While limitations remain in terms of dataset quality and spatial precision, our work sets the foundation for more robust, explainable, and farmer-accessible plant health monitoring systems. Future directions include integrating segmentation for disease localization and using LLMs to generate treatment guidance in natural language.

6. Contributions

- We proposed a fused RGB+Lab pipeline and adapted EfficientNet-B0 to accept six-channel input, improving generalization to real-world plant disease images.
- We applied targeted data augmentations (cropping, flipping, color jitter, rotation) to mitigate overfitting on the small and noisy PlantDoc dataset.
- We introduced Convolutional Block Attention Modules (CBAM) to enhance feature localization and improve classification performance in cluttered scenes.
- We conducted detailed experiments, and presented a class-wise analysis of model performance and failure cases.
- We outlined a future roadmap for integrating segmentation and natural language output to build a practical, farmer-oriented disease detection tool.

References

Dey, P., Mahmud, T., Nahar, S. R., Hossain, M. S., and Andersson, K. Plant disease detection in precision agriculture: Deep learning approaches. In *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp. 661–667. IEEE, 2024. doi: 10.1109/IDCIoT59759.2024.10467525. URL https://ieeexplore.ieee.org/document/10467525.

Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., and Pramodhini, R. Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, 3(1):305–310, 2022. doi: 10.1016/j.gltp.2022.03.016. URL https://doi.org/10.1016/j.gltp.2022.03.016.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015. doi: 10.48550/arXiv.1512.03385. URL https://arxiv.org/abs/1512.03385.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. URL https://arxiv.org/abs/1704.04861.

Schuler, J. P. S., Romani, S., Abdel-Nasser, M., Rashwan, H., and Puig, D. Color-aware two-branch dcnn for efficient plant disease classification. *MENDEL*, 28(1):55–62, 2022. doi: 10.13164/mendel.2022.1.055. URL https://doi.org/10.13164/mendel.2022.1.055.

Tan, M. and Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6105–6114. PMLR, 2019. URL https://arxiv.org/abs/1905.11946.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19. Springer, 2018. doi: 10.1007/978-3-030-01234-2_1. URL https://doi.org/10.1007/978-3-030-01234-2_1.