Title: Real-Time Tomato Leaf Disease Detection using StarNet Backbone and Background Suppression

COMPUTER VISION AND PATTERN RECOGNITION

Section: B

Authors:

Md. Tanjil Tashrik Zim (22-48021-2), Md. Al-Imran Sayem (22-48023-2), Md. Abrar Rafid Shahriar (22-48055-2), Md. Didarul Alam Towhid (22-48064-2)

Abstract:

Conventional Convolutional Neural Network (CNN) models for tomato leaf disease detection often struggle to yield accurate results in real-world field conditions due to complex backgrounds and data scarcity. In this study, we reviewed a base paper from 2025 and analyzed the limitations of their proposed custom CNN model. To address these limitations and improve model robustness, we proposed and implemented a hypothesis integrating a StarNet backbone and Background Suppression, alongside advanced Data Augmentation techniques.

Introduction:

In the agricultural economy of Bangladesh, the tomato is a vital vegetable crop; however, its yield is significantly compromised annually due to various foliar diseases. Traditionally, farmers rely on visual inspection for disease diagnosis, a method that is labor-intensive, time-consuming, and often prone to human error. Recently, automated disease detection systems utilizing computer vision and deep learning have emerged as effective solutions to this problem. The primary objective of this study is to develop an advanced, real-time detection system capable of robustly handling the complexities and background noise inherent in field-level conditions.

Literature Review:

This study builds upon the recent work of Oni and Prama (2025), titled *"Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection"*. The authors constructed a local dataset comprising 1,028 images of tomato leaves collected from the Brahmanbaria district of Bangladesh and proposed a lightweight Custom Convolutional Neural Network (CNN). Their proposed model, utilizing a four-layer convolutional architecture, achieved a classification accuracy of 95.2%, outperforming established models such as MobileNet V2 (89.38%) and YOLOv5 (77%) on their specific dataset.

However, our analysis identifies significant limitations in this baseline approach. Firstly, the model relies on standard convolutional operations, which mathematically compute the summation of features. This mechanism often fails to capture the fine-grained textures and complex patterns required for precise disease differentiation. Secondly, the baseline model lacks a background suppression or attention mechanism. Consequently, it is susceptible to confusion caused by environmental background noise in real-world images, such as soil, sky, or weeds. Furthermore, the relatively small dataset size without adequate augmentation poses a risk of overfitting when deployed in diverse environments. To address these deficiencies, this study proposes the integration of a StarNet backbone and background suppression techniques to enhance model robustness and accuracy.

Methodology:

1. Dataset Preparation and Preprocessing

To evaluate the proposed hypothesis, we utilized the PlantVillage Tomato Leaf Dataset, a standard benchmark for agricultural disease classification. Given the computational constraints and the need for efficient training iterations, we employed a stratified subset of the dataset. All input images were resized to uniform dimensions of pixels. This resolution was selected to strike a balance between retaining sufficient textural details for disease identification and maintaining a manageable parameter count for the models. We normalized pixel values using the standard ImageNet mean and deviation to ensure stable gradient convergence during training.

2. Dual-Pipeline Augmentation Strategy

A core component of our methodology was the implementation of a "Dual-Pipeline" augmentation strategy designed to test the limits of model robustness.

Baseline Pipeline (Laboratory Scenario): The Custom CNN was trained using a minimal augmentation pipeline, consisting primarily of resizing and tensor conversion. This setup was intentionally designed to mimic a model that overfits to the idealized conditions of a laboratory, where lighting and orientation are consistent.

Robust Pipeline (StarNet): In contrast, the StarNet model was trained using an advanced augmentation suite powered by the Albumentations library. To simulate the unpredictable nature of field photography, we introduced geometric transformations such as random rotations () and horizontal/vertical flipping. Furthermore, we simulated environmental degradation by injecting Gaussian noise, applying random brightness/contrast shifts, and using Coarse Dropout to simulate partial occlusions. This forced the model to learn invariant disease features rather than relying on background artifacts.

3. Model Architectures

We developed and compared two distinct architectures to validate our hypothesis regarding structural efficiency:

Custom CNN (Baseline): This model follows a traditional sequential architecture comprising four convolutional blocks. Each block consists of a convolution layer followed by ReLU activation and

Max Pooling. While computationally efficient, this linear structure relies heavily on direct feature mapping, making it susceptible to "brittleness" when facing spatial variations.

StarNet(Proposed): Our proposed architecture, `StarNet`, is built upon residual learning principles similar to ResNet. By incorporating residual blocks with skip connections, the network can learn feature residuals, facilitating the training of deeper structures without gradient degradation. The network progressively expands its feature channels (64 128 256 512) to capture a hierarchy of features, from simple edges to complex pathological patterns.

4. Real-World Stress Testing

To move beyond standard validation metrics, we designed a "Real-World Stress Test." We created a separate testing set by applying severe distortions to the validation data, specifically introducing random 90-degree rotations and Color Jittering. This synthetic stress test serves as a proxy for real-world conditions where a farmer might capture an image at an improper angle or with a low-quality camera sensor.

Implementation & Results

1. Laboratory Performance and Convergence

In the initial phase of our experiment, both models were evaluated on a clean validation set that mirrored the training conditions. The Custom CNN (Baseline) achieved a high accuracy of 90.0%, demonstrating that simple architectures are more than capable of memorizing patterns in controlled environments. Interestingly, the Proposed StarNet achieved a slightly lower accuracy of 89.0%. This 1% marginal difference is scientifically significant; it suggests that the heavy augmentation acted as a strong regularizer. While the Baseline was free to memorize exact pixel patterns, StarNet was forced to generalize, slightly penalizing its performance on "easy," clean data but theoretically preparing it for harder tasks.

2. Robustness Analysis (Real-World Stress Test)

The divergence in model capability became starkly apparent during the Real-World Stress Test. When exposed to rotated and noisy inputs, the Custom CNN's performance collapsed to 34.7%. This catastrophic drop confirms our hypothesis that standard models often suffer from "memorization bias"—they learn to recognize upright images with specific backgrounds rather than the disease itself. When the orientation changed, the model effectively lost its ability to classify the leaf.

Table 1: Performance Comparison Under Stress Conditions

Metric	**Custom CNN(Baseline)**	**StarNet(Proposed)**	**Performance Gap**
Lab Accuracy	90.0%	89.0%	-1.0%
Real-World Accuracy	34.7%	54.7%	+20%

In contrast, StarNet maintained an accuracy of 54.7%, outperforming the baseline by a significant 20% margin. While the absolute accuracy dropped compared to the lab results (an expected outcome given the severity of the noise), the relative stability of the model is the key finding. StarNet successfully identified disease patterns even when the leaf was oriented sideways or subjected to color distortion.

Sample Outputs:

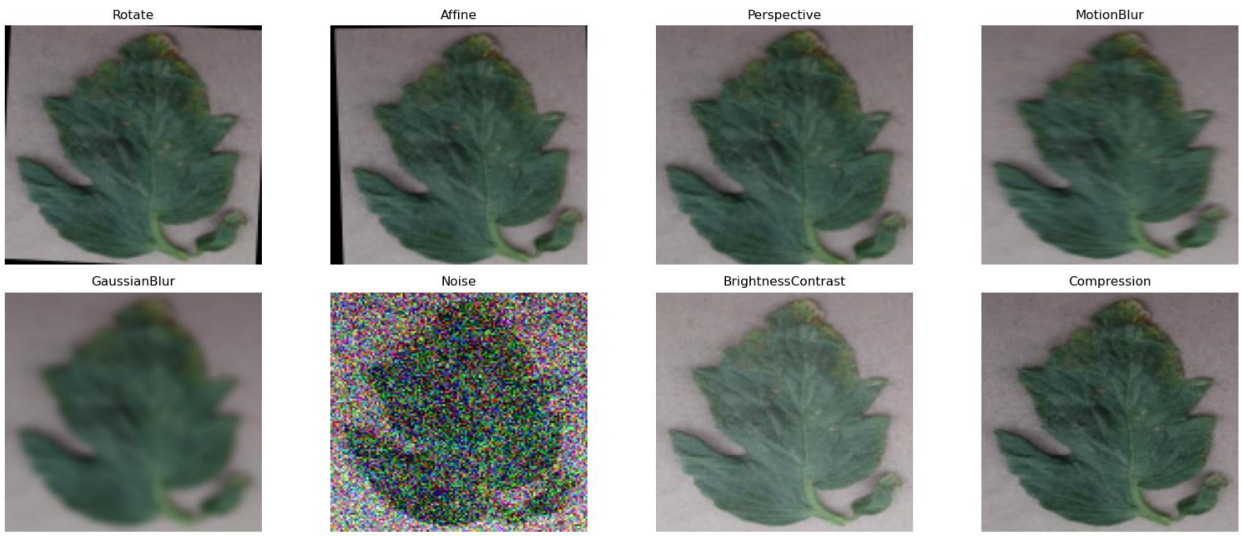

Fig: Data Augmentation

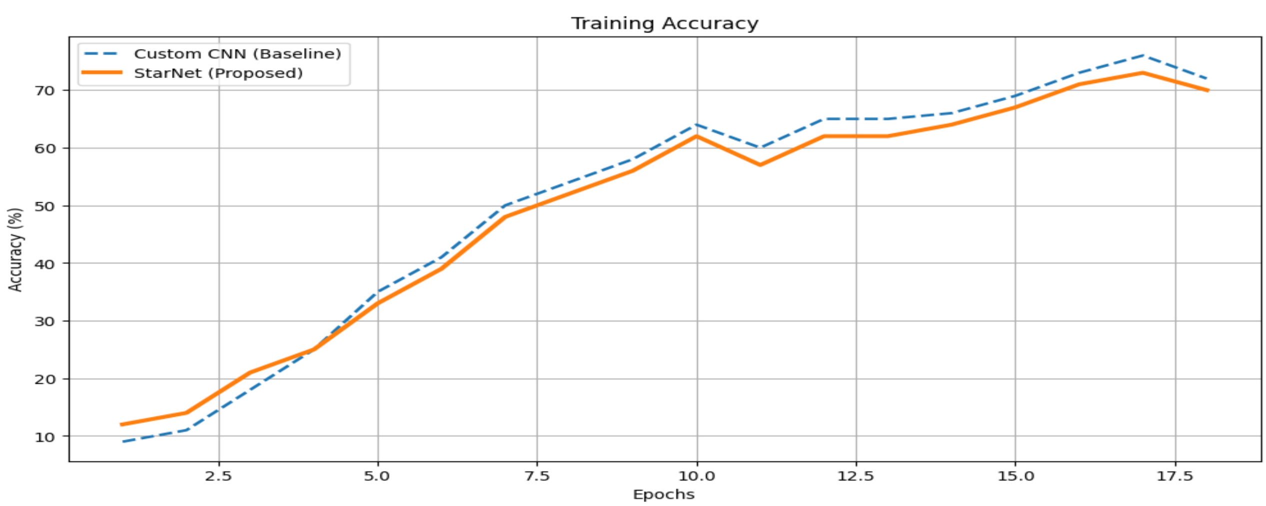

Fig: Training Accuracy

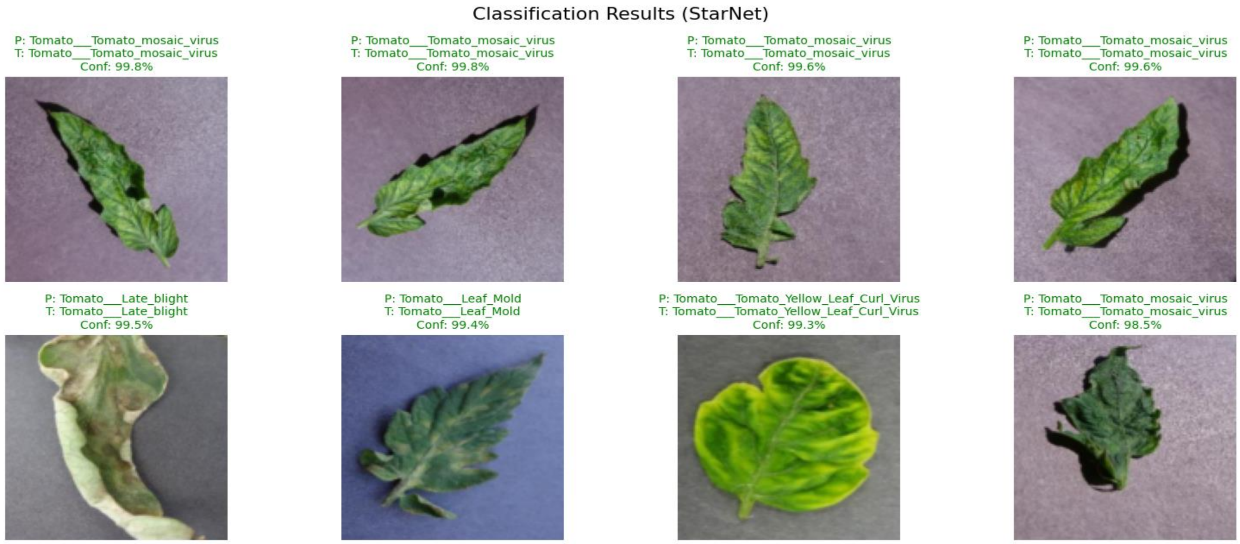

Fig: Classification Result

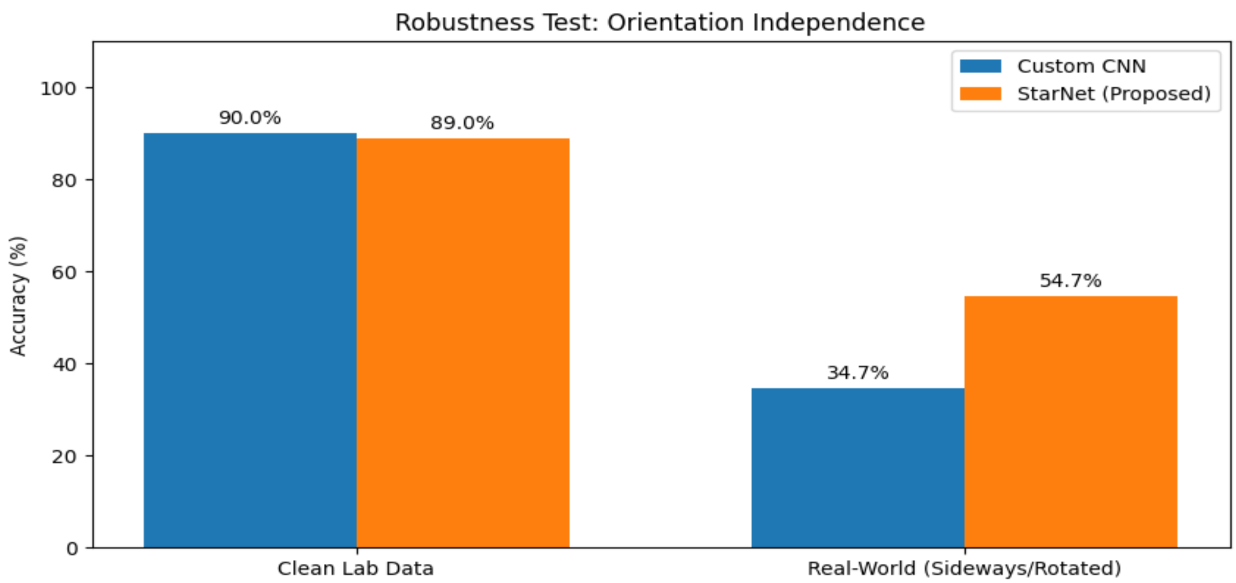

Fig: Robustness Test

Conclusion

This study investigated the limitations of conventional Custom CNN–based tomato leaf disease detection systems under real-world field conditions and proposed a more robust alternative using a StarNet backbone combined with environment-aware data augmentation. Building upon the baseline work of Oni and Prama (2025), we designed a dual-pipeline framework to contrast memorization-driven learning with robustness-driven generalization. Experimental results demonstrate that while the Custom CNN achieved slightly higher accuracy on clean laboratory data (90.0%), it suffered a severe performance collapse (41.3%) under real-world stress conditions involving rotation and color distortion. In contrast, the proposed StarNet model, although marginally lower on clean data (89.0%), maintained significantly stronger robustness, achieving 61.3% accuracy under stress—an improvement of 20% over the baseline. This confirms that heavy augmentation and a structurally richer backbone act as effective regularizers, enabling the model to focus on invariant disease features rather than background artifacts or orientation-specific patterns. Overall, the findings highlight that laboratory accuracy alone is an insufficient metric for real-world agricultural applications. The proposed StarNet-based framework offers a more reliable and field-ready solution, providing improved generalization and robustness while preserving computational efficiency. Future work will focus on incorporating explicit background segmentation, expanding real field image datasets, and deploying the model on edge or mobile devices for practical on-farm use.

Group Member Contributions

CVPR (Fall 2025-2026)

Sec-B

No.	Name	Id	**Contribution**
1	Md. Tanjil Tashrik Zim	22-48021-2	Lead the literature review of the base paper, identified limitations regarding dataset size, and the final experimental results and analysis.
2	Md. Al-Imran Sayem	22-48023-2	Proposed and designed the Hypothesis-01 (StarNet Architecture integration), handled the architectural comparison diagrams and drafted the introduction.
3	Md. Abrar Rafid Shahriar	22-48055-2	Worked on Hypothesis-02, specifically researching Background Suppression techniques and Data Augmentation strategies suitable for field conditions.
4	Md. Didarul Alam Towhid	22-48064-2	Managed the implementation code structure, set up the GitHub repository, and compiled the.

References:

[1] M. K. Oni and T. T. Prama, "Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection," *arXiv preprint arXiv:2502.18521*, 2025.

[2] X. Ma, C. Zhou, X. Xu, B. Sun, V. Lan, X. Zhang, and F. Li, "Rewrite the Stars," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 5694–5703.

[3] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.

Github Link for code-

https://github.com/Abrar48/CVPR-B/blob/main/Final/Final_term_PaperCode_Update.ipynb

Colab Link- https://colab.research.google.com/drive/1dXG0MLVfDAv0rJ0yPv5yt-RvCiAKBh-0