EAST WEST UNIVERSITY

Project Report on Task 2

Title: Analysis of Tomato Leaf Diseases using PlantVillage Dataset

Course Title: Artificial Intelligence

Course Code: CSE366

Semester: Spring 2026

Section: 01

Submitted by:

MD. Rejaul Hoq	**2023-1-60-124**
Nahidul Islam	**2022-3-60-072**
Md. Rahat Khan	**2021-1-60-090**
Meghla Biswas	**2023-1-60-225**

Submitted to

Dr. Raihan Ul Islam

Associate Professor, Department of Computer Science and Engineering

Date of Submission: 20/02/2026

Related Work

Title	Dataset name	Dataset description	Method	Accuracy	Research Questions	Pros & Cons	Citation
Tomato Leaf Disease Detection using Generative Adversarial Network-based ResNet50V2	PlantVillage Dataset https://www.kaggle .com/datasets/emmarex/plantdisease	The PlantVillage dataset containing 16,012 tomato leaf images categorized in ten classes. Among them, nine classes are tomato leaf diseases and one class are healthy leaf images with images of size 256×256 pixels	GAN-based Data Augmentation with Transfer Learning using ResNet50V2	Accuracy: 99.75% Precision: 99.28% Recall: 99.43% F1-score: 99.67%	Can Generative Adversarial Network (GAN)-based data augmentation improve the classification accuracy and generalization of tomato leaf disease detection using deep CNN models?	Pros: • Reduces overfitting using synthetic data • Achieves very high classification accuracy • Improves generalization performance Cons: • GAN training increases computational cost • Overall training time is higher	[1]

Tomato Leaf Disease Detection using Deep Learning Techniques	Custom Tomato Leaf Dataset (No public URL provided; laboratory-acquired dataset)	The dataset consists of 735 tomato leaf images belonging to 7 classes, including six disease categories and one healthy class. The dataset was split into 70% training and 30% testing samples.	Fuzzy Support Vector Machine (Fuzzy-SVM), Convolutional Neural Network (CNN), and Region-based Convolutional Neural Network (R-CNN)	96.735% (R-CNN)	Which machine learning or deep learning classifier (Fuzzy-SVM, CNN, or R-CNN) provides the highest accuracy for tomato leaf disease detection?	Pros: • R-CNN achieved higher accuracy compared to CNN and Fuzzy-SVM • Effective feature extraction using deep learning • Suitable for early disease detection Cons: • Small dataset size • High computational cost for R-CNN • Dataset collected in controlled laboratory conditions only	[2]
ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network	PlantVillage Dataset — https://www.kaggle.com/datasets/emmarex/plantdisease	Tomato leaf images collected from PlantVillage dataset. Total 10 classes (9 disease + 1 healthy). Training	Custom Convolutional Neural Network (CNN) with 3 convolution layers, 3 max-pooling layers, and 2 fully	Average classification accuracy of 91.2% across 10 classes	Can a lightweight CNN architecture outperform pre-trained models (VGG16, Inception V3, MobileNet) for	**Pros:** High accuracy with low computational cost; fewer parameters; suitable for mobile devices. **Cons:** Limited to tomato leaves;	[3]

		set: 10,000 images (1000 per class), Validation set: 7,000 images (700 per class), Test set: 500 images (50 per class). Image size: 256×256.	connected layers		tomato leaf disease classification using PlantVillage dataset?	performance may degrade with real-field images and larger datasets.	
Research on Tomato Leaf Disease Classification Based on Machine Learning	PlantVillage Dataset URL: https://github.com/spMohanty/PlantVillage-Dataset	Total 10 classes, 1 healthy tomato leaf 9 tomato leaf diseases. High-resolution tomato leaf images. Dataset taken from PlantVillage. Data augmentation used (rotation, blur, noise, color	AlexNet ResNet18 YOLOv5 YOLOv8	AlexNet: 89.25% ResNet18 : 92.28% YOLOv5: 94.45% YOLOv8: 95.18% (highest accuracy)	How effectively can machine learning and deep learning models classify tomato leaf diseases? Which deep learning model provides the best accuracy and training efficiency for tomato leaf	**Pros:** 1.High classification accuracy 2.Comparison of multiple models 3.YOLOv8 shows fast training and high performance 4.Uses standard PlantVillage dataset **Cons:** 1.Dataset has simple	[4]

		transform)			disease classification?	background images 2.Real-world complex field conditions not fully addressed 3.Limited to tomato leaves only	
Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection	Custom Field-Collected Tomato Leaf Dataset (Bangladesh) URL (dataset paper): https://doi.org/10.1016/j.dib.2025.111327	Total images: 1028, 482 healthy tomato leaf images and 546 diseased tomato leaf images. Classes: 2 (Healthy, Diseased) Images captured in real field conditions	Custom CNN (proposed), YOLOv5, MobileNetV2, ResNet18	Custom CNN: 95.2% (best) MobileNetV2: 89% YOLOv5: 84% ResNet18: 82%	1.Can a lightweight custom CNN outperform pre-trained models for tomato leaf disease detection? 2.How effective is a real-time, web-based deep learning system for assisting farmers in disease identification?	**Pros:** 1.Uses real field images, not lab-only data 2.Lightweight custom CNN (low storage, fast inference) 3.High accuracy with fewer parameters **Cons:** 1.Only binary classification (healthy vs diseased)	[5]

						2.Disease types are not separately classified	

References:

[1] A. K. Pandey, D. Jain, T. K. Gautam, J. S. Kushwah, S. Shrivastava, R. Sharma, and P. Vats,"Tomato Leaf Disease Detection using Generative Adversarial Network-based ResNet50V2," Engineering Letters, vol. 32, no. 5, pp. 965–973, May 2024.

[2] Nagamani H. S., Sarojadevi H., "Tomato Leaf Disease Detection using Deep Learning Techniques," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 13, No. 1, 2022.

[3] Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). *ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network*. Procedia Computer Science, 167, 293–301. Elsevier.

[4] Xia Yuting. *Research on Tomato Leaf Disease Classification Based on Machine Learning*. Proceedings of CIBDA 2025, ACM, DOI: 10.1145/3746709.3746883

[5] Prama, T. T., & Oni, M. K. *Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection*. Springer, ICCS 2025. DOI: 10.1007/978-3-031-97564-6_6

Model Evaluation & Comparison

We evaluated three deep learning models—ResNet50, VGG16, and EfficientNetB0—on a 10-class tomato disease classification dataset. The models were trained for up to 50 epochs with early stopping, and metrics including accuracy, precision, recall, F1-score, AUC, training time, and testing time were recorded.

Summary of Results:

Model	Accuracy	Precision	Recall	F1-score	AUC	Training Time	Testing Time
ResNet50	99.44%	99.37%	99.35%	99.36%	1.000	~42.5 min	17.22 sec
VGG16	95.47%	95.01%	95.16%	94.85%	0.9992	~90 min	28.70 sec
EfficientNetB0	99.60%	99.63%	99.53%	99.58%	1.000	~29 min	9.29 sec