https://doi.org/10.46335/IJIES.2025.10.8.1 e-ISSN: 2456-3463
Vol. 10, No. 8, 2025, PP. 1-5

International Journal of Innovations in Engineering and Science, www.ijies.net

Early Detection of Tomato Leaf Diseases Using Hybrid Machine Learning Techniques

Rita Vora [1], Sanjay Pandey[2], Ravindra Duche[3]

[1]Research Scholar,[2 3] Professor, 0009-0002-3432-5092
Department of Physics/Electronics, ISBM University, Nawapara, Chhattisgarh, India

Email od Corresponding Author: ***rita.a.vora@gmail.com***

***Received on**: 09 May,2025* ***Revised on**: 15 June,2025* ***Published on**: 17 June,2025*

Abstract – *Disease detection in tomato leaves is a critical aspect of ensuring healthy plants and maximizing agricultural productivity. In this research work, we are doing a hybrid machine learning-based analysis for true identification of healthy and diseased leaves of tomato plants. It achieves an accuracy of 92.5% with a precision of 91.8% and recall of 93.2%, indicating its efficacy to restrict false positives and false negatives. Thus, it can be a useful tool for farmers with an AUC-ROC score of 0.97. The findings underscore its ability to mitigate crop loss and improve disease management strategies. Future work will focus on improving disease detection, building the model's database, and incorporating the model into several real-time monitoring systems for a wide range of farming applications.*

Keywords- ***Tomato leaf disease detection, disease classification, Support Vector Machine (SVM), Random Forest (RF), agricultural productivity***

I. INTRODUCTION

Tomato (Solanum lycopersicum) is among the most widely grown vegetable crops in the world, with significant economic value in terms of consumption and agricultural production. Its cultivation, however, is severely affected by various diseases such as gray leaf spot, leaf curl virus, and late blight (Panno et al., 2021). These diseases not only reduce crop yield but also impact the quality of the produce, posing a risk to food security and farmer incomes. Gray leaf spot, in particular, is a significant risk to tomato production, as it thrives in humid environments with warm temperatures (Rufaye et al., 2024).Better monitoring and control measures are essential, given the economic implications of such diseases. Identification of tomato leaf diseases primarily relies on visual inspections performed by farmers, which results in delayed detection and suboptimal treatment strategies (Saeed et al., 2023). With the advent of AI and machine learning, there is significant potential for the early diagnosis of tomato leaf diseases. Transfer learning applications such as convolutional neural networks (CNNs) have been effectively used for automating disease diagnosis, allowing timely intervention (Ahmed et al., 2022; Saeed et al., 2023).CNN-based frameworks have shown promise in detecting tomato leaf diseases with accuracies above 95% (Ahmed et al., 2022; Saeed et al., 2023; Ullah et al., 2023). Lightweight models designed for mobile devices can further increase accessibility in rural areas (Bhujel et al., 2022). The role of environmental factors such as moisture and temperature also highlights the need for dynamic prediction models (Lee, 2022). Combining AI with environmental monitoring facilitates predictive systems that help farmers in proactive decision-making (Jena & K, 2022; Liu, 2023).This paper presents a comprehensive analysis of tomato leaf disease

https://doi.org/10.46335/IJIES.2025.10.8.1 e-ISSN: 2456-3463
Vol. 10, No. 8, 2025, PP. 1-5

prediction methodologies, aiming to enhance tomato crop management strategies through predictive analytics.

II. LITERATURE REVIEW

The development of deep learning methodologies in agriculture has accelerated significantly, with tomatoes being a primary focus. Several recent studies demonstrate the successful application of CNNs for tomato disease classification.Madupuri et al. (2023) used pre-trained CNNs across various climatic and soil conditions, showing robust prediction capabilities. Chowdhury et al. (2021) employed deep feature extractors like VGG16, ResNet50, and ResNet152, achieving 83% precision with VGG16. Bhandari et al. (2023) utilized EfficientNetB5 and achieved a training accuracy of approximately 99.84%, even without segmented images.Xu et al. (2024) developed an image-text retrieval method using LAFANet, optimizing diagnostic processes by integrating visual and textual data. Wu et al. (2024) proposed TMT-YOLOv5s, a YOLO-based method with integrated attention modules, improving precision in complex environments. Upadhyay (2024) enhanced ResNet-50 to address real-time detection challenges. Hossain et al. (2023) introduced transformer-based architectures, combining attention mechanisms with CNNs for versatile disease diagnosis. Liu (2023) introduced federated learning to ensure privacy and reliability in wireless settings. Buchke and Mayuri (2023) demonstrated transfer learning with Efficient Net across datasets, highlighting scalability. Xin and Li (2024) proposed machine learning-based detection systems effective across environmental contexts. These advancements collectively aim to improve sustainability, monitoring capabilities, and the protection of tomato crop yield and quality.

III. METHODOLOGY

A hybrid approach that combines with machine learning models such as Support Vector Machine (SVM) and Random Forest (RF) can effectively detect tomato leaf diseases.

Given a dataset of tomato leaf images $X = \{x1, x2, \ldots, xn\}$ with corresponding labels

$Y = \{y1, y2, \ldots, yn\}$, , the goal is to develop a machine-learning model $f: X \to Y$

a. Dataset Preparation and Preprocessing

The dataset for this study was primarily sourced from the PlantVillage dataset, a widely recognized and publicly available repository of plant disease images. It included 10,000 tomato leaf images covering healthy leaves and diseases such as Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, and Bacterial Spot. Images were resized to 224x224 pixels, normalized, and augmented using techniques like rotation, flipping, zoom, and brightness adjustments to enhance diversity and prevent overfitting. The dataset was split into training (70%), validation (15%), and testing (15%) sets. Noise reduction methods like Gaussian blur were applied, and RGB channels were extracted for improved feature learning. This preprocessing ensured a balanced, high-quality dataset, enabling the model to generalize effectively and achieve robust performance in tomato leaf disease detection.

b. Feature Extraction Methodologies

In this research work, the following Feature extraction techniques are implemented

i. Color Features (Histogram & Color Moments)

Color Features (Histogram & Color Moments): Identify disease-related color changes (Zhang et al., 2019; Picon et al., 2019).[20,21].Histogram Features (RGB, HSV, Lab spaces)

$$H_c = \sum_{i=1}^{N} p_i \qquad (1)$$

where p_i is the pixel intensity of color channel c, and N is the total number of pixels.

- Color Moments (Mean, Standard Deviation, Skewness):

$$\text{Mean:}\mu_c = \frac{1}{N}\sum_{i=1}^{N} p_i \qquad (2)$$

$$\text{Standard Deviation:}\sigma_c = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \mu_c)^2}$$

$$\text{Skewness: } S_c = \frac{1}{N}\sum_{i=1}^{N}(\sigma_c p_i - \mu_c)^3$$

ii. Texture Features (GLCM & LBP)

- **Gray-Level Co-occurrence Matrix (GLCM)**

Texture Features (GLCM & LBP): GLCM captures spatial intensity relationships (Kamilaris & Prenafeta-Boldú, 2018; Too et al., 2019), while LBP extracts texture patterns (Liu et al., 2018).GLCM represents the spatial relationship between pixel intensities at a given offset (d, θ). The matrix element P(i,j) is defined and from GLCM, statistical texture descriptors such as contrast, correlation, energy, and homogeneity were computed[22,23]:

Contrast $C = \sum_{i,j} P(i,j)(i-j)^2$

Correlation:$R = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}$

Energy: $E = \sum_{i,j} P(i,j)^{\ 2}$

Homogeneity:$H = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$

- **Local Binary Pattern (LBP):**

LBP extracts texture patterns from leaf images by converting pixel neighborhoods into binary codes [24]

$$LBP_{p,r} = \sum_{i=0}^{p-1} s(g_i - g_c)2^i \quad (3)$$

c. Classification Using SVM and Random Forest

SVM is used for high-dimensional data, employing the RBF kernel for feature space mapping (Ferentinos, 2018; Zhang et al., 2020). RF uses multiple decision trees and majority voting.Ensemble Model for Final Prediction.SVM and RF predictions are combined via ensemble learning to improve accuracy.**Support Vector Machine (SVM)** is used as the classifier due to its robustness in handling high-dimensional data. Given training data (xi, yi) where $x_i \in R^n$ and $y_i \in \{-1,1\}$ are the labels, the SVM classifier finds the optimal hyperplane:[28,29]

$$f(x) = wTx + b \quad (4)$$

where w is the weight vector, and b is the bias. The optimal hyperplane is obtained by solving

$$\min_{w,b} \frac{1}{2}|w|^2$$

subject to: $[y_i(w^T x_i + b) \geq 1, \quad \forall i]$

The RBF kernel is used to map input features into a higher-dimensional space

$$[K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2)] \quad (5)$$

Random Forest (RF): RF constructs multiple decision trees and performs majority voting:

$$P(y|Z) = \frac{1}{T}\sum_{t=1}^{T} h_t(Z) \quad (6)$$

Ensemble Model for Final Prediction

To improve accuracy, SVM and RF predictions are combined using an ensemble approach:

$$\hat{t} = \arg\max P\ (y|Z) \quad (7)$$

where: $P(y \mid Z) = \alpha \cdot PSVM(y \mid Z) + \beta \cdot PRF(y \mid Z)$

where $\alpha + \beta = 1$ are weight factors.

IV. RESULT AND DISCUSSION

Table 1 shows that the tomato leaf disease detection proposed model performed well on different evaluation metrics. The model that was developed to detect tomato leaf disease achieved high performance with an accuracy of 92.5%, precision of 91.8%, and recall of 93.2%, proving to be reliable in performing classification between healthy and infected leaves. It is important for preventing crop losses due to disease, the low false negative rate (6.8%) ensures that diseases are detected at an early stage, while the low false positive (6.0%) reduces unnecessary rubbing. Evaluation Results — The AUC-ROC score is 0.97, indicating effectiveness in distinguishing classes.

The high performance is due to the use of balanced dataset and hybrid machine learning techniques. Nevertheless, the false negative rate of 6.8% indicates that we can do better, especially when it comes to detecting diseases at an early stage. Next steps could be to use higher resolution images and the integration of multi-spectral data, as well as expanding the dataset to include rare diseases and environmental variations.

A model like this has massive real world applications, helping farmers monitor crop health remotely and making data-driven decisions. Its rapid inference time makes it paper little to no real-time use, and it can be incorporated into drones or even mobile devices. The model provides a reliable solution for tomato leaf disease detection and could significantly contribute towards innovations in agriculture.

Table 1: Performance Matrix for Tomato Leaf Disease Detection

Metric	Value (%)
Accuracy	92.5
Precision	91.8
Recall	93.2
F1-Score	92.5
Specificity	94.0
False Positive Rate (FPR)	6.0
False Negative Rate (FNR)	6.8
AUC-ROC	0.97

V. CONCLUSION AND FUTURE SCOPE

The accuracy, precision, and recall obtained by the proposed model where accuracy obtained is 92.5%, precision obtained is 91.8% while the recall achieved is 93.2% indicating the effectiveness of the proposed model to identify healthy and disease status of leaves. It has a low false positive (6.0%) and negative (6.8%) rate which ensures reliable detection of disease and assists farmers in early management of crops. Moreover, the AUC-ROC score of 0.97 demonstrates the robustness of the model classification. We believe that these results strongly show how the model has the potential to minimize crop losses, and maximize yield in agriculture. Future works may include enhancing detection in early stage diseases, extending database coverage for rare diseases and environmental variations, and translating the model to drones or mobile apps for real time monitoring. The model could also be expanded to other crops, and the performance could be increased by enhancing interpretability, optimizing for edge devices, etc.

REFERENCES

[1] *Ahmed, S., Hasan, M., Ahmed, T., Sony, M., & Kabir, M. (2022). Less is more: lighter and faster deep neural architecture for tomato leaf disease classification. IEEE Access, 10, 68868–68884. https://doi.org/10.1109/access.2022.3187203*

[2] *Bhujel, A., Kim, N., Arulmozhi, E., Basak, J., & Kim, H. (2022). A lightweight attention-based convolutional neural networks for tomato leaf disease classification. Agriculture, 12(2), 228. https://doi.org/10.3390/agriculture12020228*

[3] *Jena, K., & K, K. (2022). A machine intelligent based approach for the classification and analysis of tomato leaf disease images. International Journal of Health Sciences and Pharmacy, 1–19. https://doi.org/10.47992/ijhsp.2581.6411.0083*

[4] *Lee, M. (2022). Forecasting leaf mold and gray leaf spot incidence in tomato and fungicide spray scheduling. Journal of Bio-Environment Control, 31(4), 376–383. https://doi.org/10.12791/ksbec.2022.31.4.376*

[5] *Liu, P. (2023). Tomato leaf disease detection based on peer-to-peer federated learning in wireless networks. https://doi.org/10.1117/12.2685057*

[6] *Panno, S., et al. (2021). A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the Mediterranean basin. Agronomy, 11(11), 2188. https://doi.org/10.3390/agronomy11112188*

[7] *Rufaye, Z., Sheikh, W., & Ghaleb, I. (2024). Detection of sm gene resistance to gray leaf spot in (Stemphylium spp.) tomato cultivars Iraq. Comunicata Scientiae, 15, e3564. https://doi.org/10.14295/cs.v15.3564*

[8] *Saeed, A., et al. (2023). Smart detection of tomato leaf diseases using transfer learning-based convolutional neural networks. Agriculture, 13(1), 139. https://doi.org/10.3390/agriculture13010139*

[9] *Ullah, N., et al. (2023). A lightweight deep learning-based model for tomato leaf disease classification. Computers, Materials & Continua, 77(3), 3969–3992. https://doi.org/10.32604/cmc.2023.041819*

[10] *Bhandari, M., et al. (2023). Botanicx-AI: Identification of tomato leaf diseases using an explanation-driven deep-learning model. Journal of Imaging, 9(2), 53. https://doi.org/10.3390/jimaging9020053*

[11] *Buchke, P., & Mayuri, V. (2023). Recognize and classify illnesses on tomato leaves using EfficientNet's transfer learning approach with different size dataset. https://doi.org/10.21203/rs.3.rs-3149045/v1*

[12] *Chowdhury, M., et al. (2021). Tomato leaf diseases detection using deep learning technique. https://doi.org/10.5772/intechopen.97319*

[13] *Hossain, S., et al. (2023). Aggregating different scales of attention on feature variants for tomato leaf disease diagnosis: A transformer-driven study. Sensors, 23(7), 3751. https://doi.org/10.3390/s23073751*

[14] *Madupuri, R., et al. (2023). Deep learning image-based automated application on classification of tomato leaf disease by pre-trained CNNs. Mehran University Research Journal of Engineering and Technology, 42(3), 52. https://doi.org/10.22581/muet1982.2303.06*

[15] *Upadhyay, A. (2024). Evaluation of enhanced ResNet-50 based deep learning classifier for tomato leaf disease detection and classification. JES, 20(3s), 2270–2282. https://doi.org/10.52783/jes.1856*

[16] *Wu, Y., et al. (2024). A lightweight tomato leaf disease detection system with strong generalization based on TMT-YOLOv5s. Academic Journal of Computing & Information Science, 7(2). https://doi.org/10.25236/ajcis.2024.070210*

[17] *Xin, D., & Li, T. (2024). Revolutionizing tomato disease detection in complex environments. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1409544*

[18] *Xu, J., et al. (2024). High-accuracy tomato leaf disease image-text retrieval using LAFANet. Plants, 13(9), 1176. https://doi.org/10.3390/plants13091176*

[19] *Zhang, S., et al. (2019). Plant disease leaf image segmentation based on color features and clustering algorithms. Computers and Electronics in Agriculture, 165,104934. https://doi.org/10.1016/j.compag.2019.104934*

[20] *Picon, A., et al. (2019). Crop-conditional CNNs for massive multi-crop disease classification via smartphone images. Computers and Electronics in Agriculture, 167, 105093. https://doi.org/10.1016/j.compag.2019.105093*

[21] *Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016*

[22] *Too, E. C., et al. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture, 161, 272–279. https://doi.org/10.1016/j.compag.2018.03.032*

[23] *Liu, B., et al. (2018). Identification of apple leaf diseases using deep CNNs. Symmetry, 10(1), 11. https://doi.org/10.3390/sym10010011*

[24] *Zhang, X., et al. (2020). Identification of maize leaf diseases using improved CNNs. IEEE Access, 8, 30352–30361. https://doi.org/10.1109/ACCESS.2020.2972412*

[25] *Saleem, M. H., et al. (2019). Plant disease detection and classification by deep learning. Plants, 8(11), 468. https://doi.org/10.3390/plants8110468*

[26] *Arsenovic, M., et al. (2019). Solving limitations of deep learning-based approaches for plant disease detection. Symmetry, 11(7), 939. https://doi.org/10.3390/sym11070939*

[27] *Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009*

[28] *Zhang, S., Huang, W., & Zhang, C. (2019). Three-channel CNNs for vegetable leaf disease recognition. Cognitive Systems Research, 53, 31–41. https://doi.org/10.1016/j.cogsys.2018.04.006*