Deep Learning for Tomato Disease Detection and Severity Assessment: A Systematic Analytical Review of Methods, Datasets, and Challenges

Shubham Sandal
shubhamsandal551@gmail.com

Lovely Professional University

Sukanta Ghosh
Lovely Professional University

Research Article

Keywords: Tomato disease detection, Deep learning, Disease severity assessment, Image segmentation, Vision Transformers, Precision agriculture

Posted Date: May 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9600064/v1

Additional Declarations: No competing interests reported.

Title: Deep Learning for Tomato Disease Detection and Severity Assessment: A Systematic Analytical Review of Methods, Datasets, and Challenges

Authors:

Shubham Sandal [1*], Dr. Sukanta Ghosh [2]

Affiliations:

[1] Research Scholar, Lovely Professional University, Phagwara, Punjab, India

[2] Associate Professor, Lovely Professional University, Phagwara, Punjab, India

Corresponding Author:

Shubham Sandal

Email: shubhamsandal551@gmail.com

Mobile: +91 9876199215

Abstract:

Tomato diseases significantly affect crop productivity and food security, necessitating accurate and timely detection methods. This paper presents a systematic and analytical review of deep learning approaches for tomato disease detection and severity assessment, based on 76 research studies. Existing methods are categorized into classification, detection, segmentation, and emerging multi-task frameworks. The analysis shows that convolutional neural networks achieve high accuracy on controlled datasets but exhibit limited generalization in real-world conditions. Advanced architectures, including transformer-based and hybrid models, improve performance but increase computational complexity. A key finding is the limited focus on disease severity assessment, which remains underexplored despite its importance for precision agriculture. The review identifies major challenges, including dataset limitations, lack of standardized benchmarks, and deployment constraints. Future directions emphasize multi-task learning, real-world dataset development, lightweight models, and explainable AI. This study provides a foundation for developing robust and practical tomato disease detection systems.

Keywords: Tomato disease detection; Deep learning; Disease severity assessment; Image segmentation; Vision Transformers; Precision agriculture

1. Introduction

1.1 Importance of Tomato Disease Detection

Tomato is one of the most widely cultivated and economically significant crops globally, contributing substantially to food security and agricultural economies [1], [2]. However, tomato crops are highly susceptible to a wide range of diseases and pests, which can significantly reduce yield and quality, sometimes leading to severe economic losses [3], [4]. Early and accurate detection of these diseases is therefore critical to ensuring crop productivity and sustainable agricultural practices [1].

Traditional disease identification methods rely heavily on human expertise and manual inspection, which are time-consuming, labor-intensive, and prone to errors [5], [6]. Moreover, delayed or inaccurate diagnosis can result in improper treatment decisions, leading to further spread of diseases and increased crop damage [7], [8]. In addition, global challenges such as climate change and increasing food demand further emphasize the need for efficient and automated disease detection systems [9], [10].

1.2 Role of Deep Learning

Deep learning (DL) has emerged as a transformative technology in plant disease detection due to its ability to automatically extract complex features from images without manual intervention [5], [11], [12]. Unlike traditional machine learning approaches that rely on handcrafted features, DL models such as Convolutional Neural Networks (CNNs) can learn hierarchical representations directly from raw image data, enabling more accurate and robust disease identification [13], [14], [15].

Recent advancements in DL architectures, including attention mechanisms, transformers, and multimodal frameworks, have further enhanced the capability of automated systems to detect subtle disease patterns and variations in real-world conditions [16], [17], [18]. Additionally, DL-based approaches support real-time detection and large-scale deployment, making them highly suitable for precision agriculture applications [3], [19].

1.3 Current Progress (Classification → Detection → Segmentation)

The evolution of plant disease detection systems has progressed through multiple stages, starting from classification-based approaches to more advanced detection and segmentation techniques. Initially, most studies focused on image classification, where models predict disease categories based on entire leaf images using CNN architectures, achieving high accuracy under controlled conditions [1], [20], [21]. However, these approaches lack spatial understanding of disease regions and are limited in real-world applications.

Subsequently, object detection methods such as YOLO-based architectures were introduced to localize disease regions and identify multiple disease instances within a single image, improving practical applicability in field environments [3]. These models enable real-time detection and provide more actionable insights compared to classification methods.

More recently, segmentation-based approaches have gained attention, focusing on pixel-level identification of diseased regions to enable precise localization and severity estimation [7], [22]. These methods allow detailed analysis of lesion areas, which is essential for disease severity assessment and decision-making in crop management. Furthermore, advanced research has extended beyond detection to include severity estimation, integrating tasks such as lesion segmentation, feature extraction, and disease progression analysis [23], [24].

1.4 Limitations of Existing Works

Despite significant progress, existing studies exhibit several critical limitations that hinder their practical deployment. One major issue is the reliance on controlled datasets, such as laboratory images with uniform backgrounds, which results in poor generalization to real-world field conditions [9], [25], [26]. Another limitation is the requirement for large annotated datasets, especially for deep learning models, which are expensive and time-consuming to create [13], [27]. In particular, pixel-level annotation required for segmentation and severity estimation is highly labor-intensive and often impractical at scale [23], [28].

Additionally, many models focus solely on classification tasks and ignore severity estimation, which is crucial for effective disease management [23], [29]. Models also struggle with complex environmental conditions such as varying illumination, occlusion, and intra-class variability, leading to reduced performance in real-world scenarios [16], [30]. Moreover, deep learning models are often computationally expensive, limiting their deployment on edge devices commonly used in agricultural fields [13], [31], [32].

1.5 Research Gap

From the above analysis, it is evident that there is a lack of comprehensive studies that systematically integrate detection, segmentation, and severity assessment within a unified framework. Most existing works focus on isolated tasks, primarily classification, without addressing the complete pipeline required for practical agricultural applications [23], [33].

Furthermore, there is insufficient emphasis on tomato-specific disease analysis, as many studies generalize across multiple crops without addressing the unique challenges associated with tomato diseases [3]. The integration of multimodal information, such as combining visual and semantic knowledge, is also still in its early stages and remains underexplored [16], [34]. Another significant gap lies in the limited focus on real-world deployment challenges, including model interpretability, computational efficiency, and robustness under varying environmental conditions [35], [30].

Sr. No.	Approach Type	Key Focus	Strength	Limitation
1	Classification (CNN) [1]	Disease identification	High accuracy in lab conditions	Poor localization
2	Detection (YOLO) [2]	Localization, classification	Real-time detection	Limited fine-grained analysis
3	Segmentation (GAN/UNet) [4]	Pixel-level detection	Enables severity estimation	Requires annotated data
4	Severity Estimation [8]	Lesion analysis	Practical disease management	Underexplored
5	Multimodal DL [7]	Visual + semantic	Improved interpretability	Early stage research

Table 1: Comparative analysis of deep learning approaches for plant disease detection and severity assessment, highlighting their strengths, limitations, and application scope.

1.6 Contribution of This Review

This review aims to provide a systematic and analytical evaluation of deep learning methods for tomato disease detection and severity assessment. Unlike existing surveys, this study focuses specifically on tomato diseases and analyzes the evolution of methodologies from classification to advanced segmentation and severity estimation approaches.

The key contributions of this review are as follows:

- A structured analysis of deep learning techniques used in tomato disease detection
- A comparative evaluation of classification, detection, and segmentation approaches
- Identification of key limitations and challenges in existing works
- Highlighting research gaps related to severity assessment and real-world deployment
- Providing future research directions for developing robust and scalable agricultural solutions

2. Review Methodology

2.1 Review Objective

The primary objective of this study is:

“To systematically review existing deep learning methods for tomato disease detection and severity assessment.”

This objective guides a structured and analytical evaluation of current methodologies, focusing on identifying technological trends, methodological advancements, and research gaps, particularly in disease severity assessment.

2.2 Research Questions

To address the stated objective, the following research questions are formulated:

RQ1: What are the predominant deep learning techniques used for tomato disease detection?

RQ2: How have methodologies evolved from classification to detection and segmentation?

RQ3: What datasets and evaluation metrics are commonly used in this domain?

RQ4: How is disease severity assessment addressed in existing studies?

RQ5: What are the key limitations of current deep learning-based approaches?

RQ6: What research gaps and future directions can be identified?

2.3 Data Sources

A comprehensive literature search was conducted using two major academic databases:

- Scopus (primary database)
- Google Scholar

Scopus was used as the primary indexing source to ensure the inclusion of high-quality, peer-reviewed journal articles, while Google Scholar was used to enhance coverage and minimize publication bias by capturing additional relevant studies.

2.4 Search Strategy

A structured keyword-based search strategy was adopted to retrieve relevant studies.

Primary Keywords:

- “Tomato disease detection”

- “Plant disease detection deep learning”
- “Tomato leaf disease classification”
- “Crop disease segmentation”
- “Disease severity assessment agriculture”

Search Strings (examples):

- (“tomato disease” AND “deep learning”)
- (“plant disease detection” AND “CNN” OR “Transformer”)
- (“tomato disease” AND “segmentation” OR “severity”)

Boolean operators (AND, OR) and database-specific filters were applied to refine the search results.

2.5 Inclusion Criteria

Studies were included based on the following criteria:

- Published between 2020 and 2026
- Focus on plant disease detection, with emphasis on tomato-related studies
- Utilization of deep learning, machine learning, or computer vision techniques
- Availability of sufficient methodological and experimental details
- Published in peer-reviewed or reputable academic sources
- Reporting quantitative performance evaluation metrics (e.g., accuracy, precision, IoU, F1-score)

2.6 Exclusion Criteria

The following studies were excluded:

- Duplicate records identified across databases
- Studies not directly relevant to plant disease detection or image-based analysis
- Papers lacking sufficient methodological transparency or experimental validation
- Studies without quantitative performance evaluation metrics
- Non-scholarly articles (e.g., blogs, editorials, short summaries)

2.7 Screening Process

The study selection process was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

The screening process involved the following stages:

1. Identification: A total of 115 records were identified through database searches (73 from Scopus and 42 from Google Scholar).
2. Deduplication: After removing duplicate records, 102 unique studies were retained for further analysis.
3. Title and Abstract Screening: A total of 18 studies were excluded based on irrelevance or insufficient alignment with the research objectives, resulting in 84 studies.
4. Eligibility Assessment: Full-text articles of the remaining 84 studies were assessed against the inclusion and exclusion criteria.
5. Final Inclusion: After excluding 8 studies due to insufficient methodological quality or lack of experimental validation, a total of 76 studies were included for qualitative and analytical review.

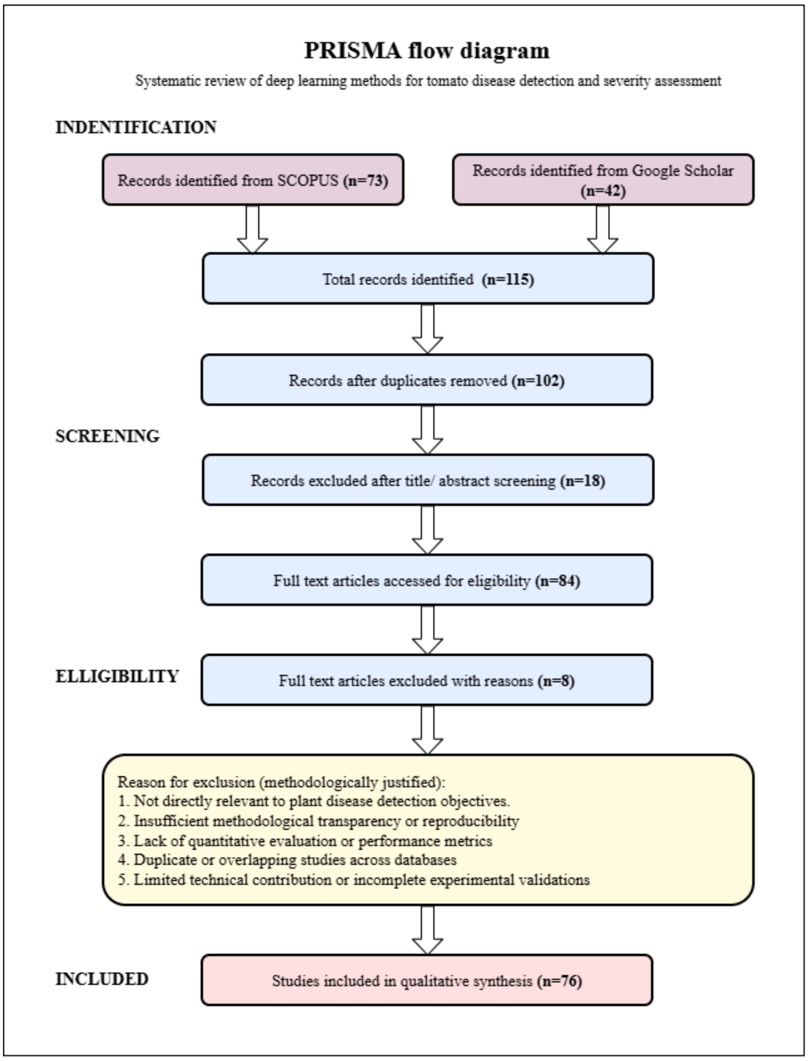

Figure 1: PRISMA Flow Diagram

2.8 Final Selected Papers

Following the PRISMA-based screening process shown in the Figure 1, a total of 76 research papers were selected for the final analysis.

These studies provide a comprehensive representation of deep learning approaches for tomato disease detection, covering classification, detection, segmentation, and emerging hybrid and multi-task learning techniques.

3. Taxonomy of Deep Learning Methods

3.1 Classification-Based Methods

Classification-based approaches form the most widely explored paradigm in tomato disease detection, where models assign an input image to predefined disease categories without explicit localization.

Deep Convolutional Neural Networks (CNNs), including DenseNet, ResNet, and EfficientNet, dominate this category due to their strong feature extraction capabilities. These architectures have demonstrated high accuracy under controlled conditions, particularly when trained on curated datasets [36]. Additionally, optimized and hybrid CNN frameworks further enhance classification performance through feature fusion and optimization techniques [37].

To improve robustness, several studies incorporate:

- Attention mechanisms for feature refinement
- Lightweight architectures for deployment efficiency
- Ensemble learning for improved generalization

Despite these advancements, classification-based models exhibit a major limitation:

Over-reliance on laboratory datasets: This results in poor generalization when applied to real-world agricultural environments characterized by complex backgrounds, varying illumination, and occlusions [38]. Consequently, classification methods, although mature, are insufficient for practical deployment without further enhancements.

3.2 Detection-Based Methods

Detection-based approaches extend classification by localizing diseased regions using bounding boxes, making them more applicable to real-world scenarios.

Modern object detection frameworks such as Faster R-CNN and YOLO enable simultaneous localization and classification. These models significantly improve interpretability and allow detection of multiple disease regions within a single image. Improved detection architectures have demonstrated enhanced accuracy and speed in identifying disease regions under field conditions [39]. Additionally, detection frameworks provide a foundation for integrating downstream tasks such as severity estimation and monitoring.

However, challenges persist:

- Difficulty in detecting early-stage or subtle disease symptoms
- Sensitivity to environmental noise
- Limited annotated datasets for detection tasks

Thus, detection methods represent an intermediate stage between classification and fine-grained analysis, offering improved practicality but still requiring further refinement.

3.3 Segmentation-Based Methods

Segmentation-based methods provide pixel-level identification of diseased regions, enabling precise delineation of infection areas.

Architectures such as U-Net and its variants are widely used due to their effectiveness in pixel-wise prediction tasks. Advanced segmentation models have demonstrated high accuracy in capturing disease-affected regions, even under complex conditions [40].

Recent developments include:

- Attention-enhanced segmentation
- Transformer-based models for global context
- Semi-supervised learning for reduced annotation dependency

Segmentation methods significantly improve:

- Spatial understanding of disease distribution
- Localization precision
- Suitability for downstream severity analysis

However, limitations include:

- High annotation cost (pixel-level labeling)

- Increased computational complexity
- Limited real-world validation

Despite these challenges, segmentation represents a critical step toward quantitative disease analysis systems.

3.4 Severity Estimation Methods

Severity estimation focuses on quantifying the extent of disease infection, which is essential for precision agriculture and decision-making. Compared to other categories, severity estimation remains relatively under-explored. Existing approaches typically rely on segmentation outputs to estimate the proportion of infected regions relative to total leaf area.

Some studies integrate lesion segmentation and classification to enable severity-related analysis, demonstrating the potential of combining spatial and semantic information [41]. However, these approaches often treat severity estimation as a secondary task rather than a primary objective.

Key limitations include:

- Lack of standardized severity metrics and datasets
- Dependence on segmentation accuracy
- Limited validation in real agricultural environments

This highlights a significant research gap, particularly in developing robust and standardized severity estimation frameworks.

3.5 Multi-Task Learning Approaches

Multi-task learning (MTL) approaches aim to simultaneously perform multiple tasks, such as classification, detection, segmentation, and severity estimation. Recent studies propose hybrid frameworks that integrate multiple deep learning architectures to improve efficiency and performance. These approaches leverage shared feature representations to reduce redundancy and enhance generalization.

For example, ensemble-based and hybrid deep learning models have demonstrated improved performance by combining multiple feature extraction strategies [42]. Additionally, multi-task frameworks enable simultaneous disease detection and severity estimation, improving system efficiency.

However, challenges include:

- Increased model complexity
- Difficulty in balancing multiple objectives
- Limited availability of multi-task annotated datasets

Despite these limitations, MTL approaches represent a promising direction for future intelligent agricultural systems.

Sr. No.	Category	Research Maturity	Strength	Limitation
1	Classification	High	High accuracy	Poor real-world generalization
2	Detection	Moderate	Localization	Weak early detection
3	Segmentation	Growing	Pixel-level precision	High annotation cost
4	Severity Estimation	Low	Decision support	No standard framework
5	Multi-task Learning	Emerging	Integrated analysis	High complexity

Table 2: Analytical taxonomy of classification, detection, segmentation, severity estimation, and multi-task learning approaches in tomato disease detection.

4. Comparative Analytical Review

This section presents a multi-dimensional analytical comparison of deep learning approaches for tomato disease detection and severity assessment.

4.1 Model Architectures: CNN vs Transformer vs Hybrid

Deep learning architectures for tomato disease detection are primarily categorized into CNN-based, Transformer-based, and Hybrid models, each with distinct strengths and limitations. CNN-based architectures dominate the literature, accounting for a significant proportion of studies (~70–80%). These models, including ResNet, DenseNet, EfficientNet, and VGG, are highly effective in extracting local spatial features and achieving high classification accuracy, particularly on structured datasets [43], [3], [44], [45]. Lightweight CNN variants further improve deployment feasibility by reducing computational overhead [46], [39].

However, CNNs inherently focus on local receptive fields, limiting their ability to capture global contextual information, which is essential for detecting diffuse or early-stage disease symptoms. These models demonstrate improved robustness under complex visual conditions such as varying illumination and background clutter [47], [36]. Despite their advantages, their adoption remains relatively limited due to high computational cost and training complexity.

Hybrid models combining CNN and Transformer components represent the most promising direction. By integrating local and global feature learning, these models achieve superior performance and improved generalization [40], [48]. Ensemble-based approaches further enhance prediction accuracy through multi-model fusion [49], [8].

Sr. No.	Model Type	Strength	Limitation	Supporting Studies
1	CNN	Efficient local feature extraction; high accuracy	Limited global context; overfitting risk	[43], [3], [44]
2	Transformer	Captures global dependencies; robust to variability	High computational cost	[47], [36]
3	Hybrid	Combines local + global features; better generalization	Increased complexity	[40], [48]

Table 3: Comparative Analysis of Model Architectures

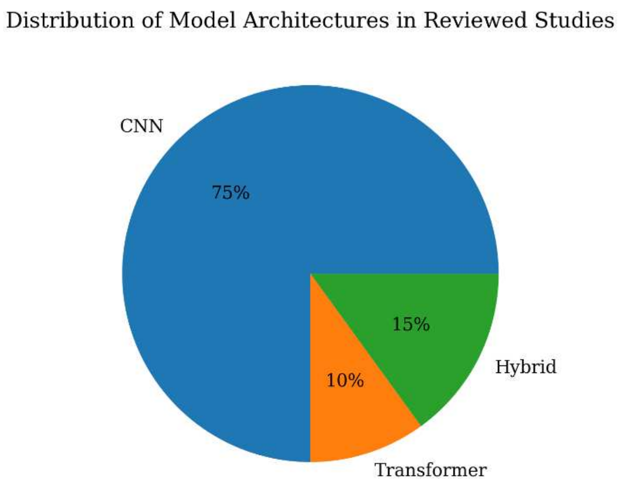

Figure 2: Architecture Comparison

As shown in table 3 and figure 2 , there is a clear architectural progression from CNN-based models toward hybrid frameworks that balance accuracy and robustness.

4.2 Dataset Analysis: PlantVillage vs Field Datasets

Dataset characteristics significantly influence model performance and generalization. A large proportion of studies rely on PlantVillage and similar laboratory datasets, which offer controlled conditions, uniform backgrounds, and minimal noise. These characteristics enable models to achieve very high accuracy (often >95%) [50], [51], [52]. However, such datasets do not represent real agricultural environments.

Field datasets introduce real-world complexities such as:

- Background clutter
- Variable lighting conditions
- Leaf occlusion and overlap

Studies utilizing field datasets consistently report lower but more realistic performance, indicating the importance of dataset diversity [53], [54], [55]. Another critical factor is dataset size and diversity. Larger and multi-source datasets improve model robustness and reduce overfitting, particularly when combining laboratory and field data [56], [57].

Sr. No.	Dataset Type	Advantage	Limitation	Supporting Studies
1	Laboratory datasets	High accuracy; controlled environment	Poor real-world generalization	[50], [51]
2	Field datasets	Realistic conditions	Annotation complexity	[53], [54]
3	Hybrid datasets	Improved robustness	Data inconsistency	[56], [57]

Table 4: Dataset-Based Comparative Analysis

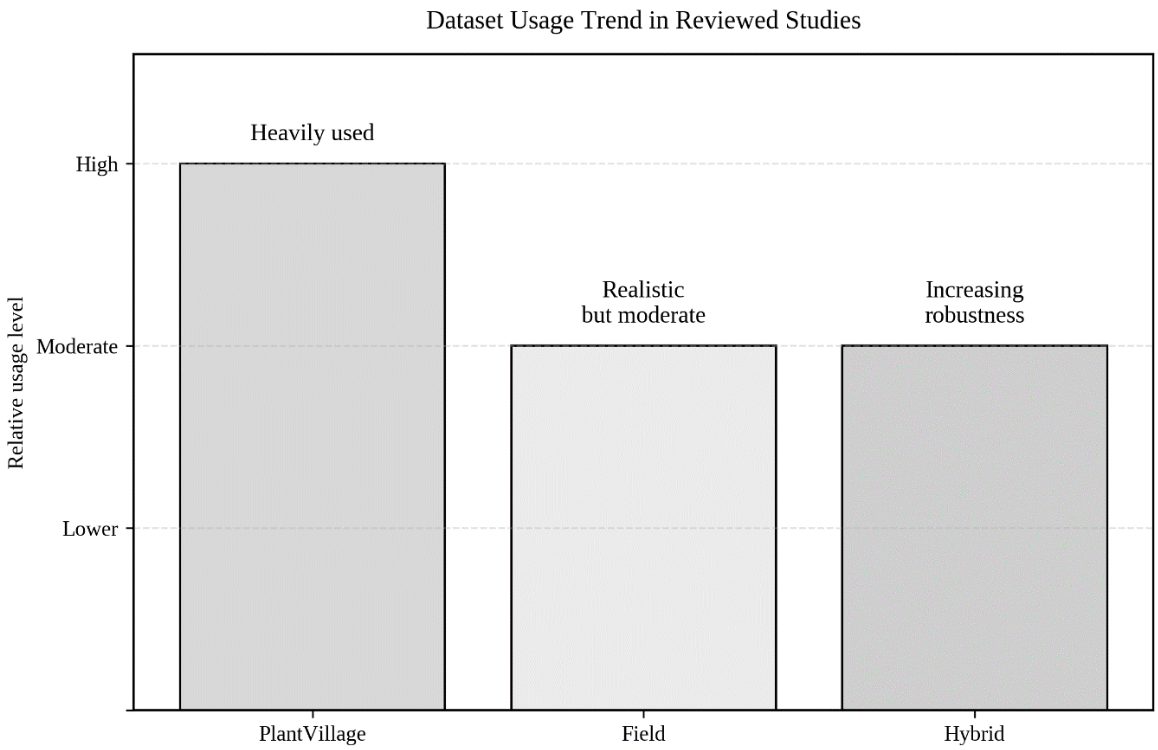

Figure 3: Dataset Comparison

Table 4 and Figure 3 highlights that dataset selection is a primary factor behind the discrepancy between reported and real-world performance.

4.3 Evaluation Metrics: Accuracy vs mAP vs mIoU

Evaluation metrics play a critical role in assessing the performance of deep learning models for tomato disease detection and severity assessment. However, the diversity of tasks—classification, detection, and segmentation—has led to the use of heterogeneous evaluation criteria, making cross-study comparison challenging [5], [13]. This section provides a detailed overview of commonly used metrics, their mathematical formulations, appropriate usage, and associated limitations.

1. Mathematical Definitions of Evaluation Metrics: Accuracy is defined as the proportion of correctly classified samples among the total number of samples:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Precision measures the proportion of correctly predicted positive instances among all predicted positives:

- Precision = TP / (TP + FP)

Recall (Sensitivity) evaluates the ability of the model to correctly identify positive instances:

- Recall = TP / (TP + FN)

The F1-score provides a harmonic mean of precision and recall, offering a balanced measure:

- F1-score = 2 × (Precision × Recall) / (Precision + Recall)

These metrics are widely used in classification-based plant disease detection studies [5], [13], [31].

For object detection tasks, mean Average Precision (mAP) is widely used. It represents the average of precision values computed at different recall levels and Intersection over Union (IoU) thresholds:

- mAP = mean(AP across classes)

Detection-based studies such as YOLO and Faster R-CNN frameworks commonly rely on mAP for performance evaluation [25], [58].

For segmentation tasks, mean Intersection over Union (mIoU) is a standard metric, defined as:

- mIoU = (Area of Overlap) / (Area of Union)

The Dice coefficient (also known as F1-score for segmentation) is defined as:

- Dice = (2 × |Prediction ∩ Ground Truth|) / (|Prediction| + |Ground Truth|)

These metrics are widely adopted in segmentation-based disease detection and severity estimation studies [59], [60], [61].

2. Task-Specific Applicability of Metrics: The choice of evaluation metric depends on the nature of the task and the type of output generated by the model.

- Classification Tasks: Accuracy and F1-score are commonly used due to their effectiveness in measuring overall prediction performance [43], [44].
- Detection Tasks: mAP is the standard metric as it evaluates both localization and classification performance simultaneously [25], [58].
- Segmentation Tasks: mIoU and Dice coefficient are preferred, as they provide pixel-level evaluation of predicted regions against ground truth annotations [60].

Thus, selecting appropriate metrics is essential for meaningful evaluation and comparison across different methodological paradigms.

3. Critical Analysis of Evaluation Metrics: Despite their widespread use, evaluation metrics exhibit several limitations that can lead to misleading interpretations.

First, accuracy can be misleading, particularly in imbalanced datasets. In plant disease detection, datasets such as PlantVillage often lead to inflated performance due to controlled conditions, whereas real-world datasets show significantly lower accuracy [50], [53]. In such cases, a model may achieve high accuracy without effectively learning disease-specific features.

Second, mAP is fundamentally different from accuracy and should not be directly compared with classification metrics. While accuracy evaluates overall correctness, mAP incorporates spatial localization through IoU thresholds and evaluates detection performance across varying confidence levels [25], [58]. Therefore, comparing mAP with accuracy across studies can lead to incorrect conclusions.

Third, segmentation metrics such as mIoU and Dice coefficient are highly dependent on the quality of ground truth annotations. Pixel-level labeling is labor-intensive and often inconsistent, which can affect the reliability of segmentation performance evaluation [23], [61].

Furthermore, the lack of standardized evaluation protocols across studies introduces additional challenges. Variations in datasets, preprocessing techniques, and metric selection make it difficult to perform fair and direct comparisons between models [36], [62].

Overall, this analysis highlights the need for standardized evaluation frameworks and the adoption of task-appropriate metrics to ensure reliable and meaningful performance assessment in tomato disease detection systems.

Evaluation metrics vary significantly across studies, depending on the task type, leading to inconsistencies in performance comparison. Classification-based studies predominantly use accuracy, precision, recall, and F1-score [43], [44]. Detection-based approaches employ mean Average Precision (mAP), while segmentation-based methods rely on mIoU, Dice score, and pixel accuracy [39], [58], [60], [61]. A major issue identified across the literature is the lack of standardized evaluation protocols, making cross-study comparisons difficult [36], [63].

Furthermore, accuracy-based evaluation often overestimates model performance, particularly when used with controlled datasets.

Sr. No.	Task Type	Metrics Used	Key Issue
1	Classification	Accuracy, F1-score	Inflated results on lab data
2	Detection	mAP	Not directly comparable with accuracy
3	Segmentation	mIoU, Dice	Dependent on annotation quality

Table 5: Evaluation Metrics Across Tasks

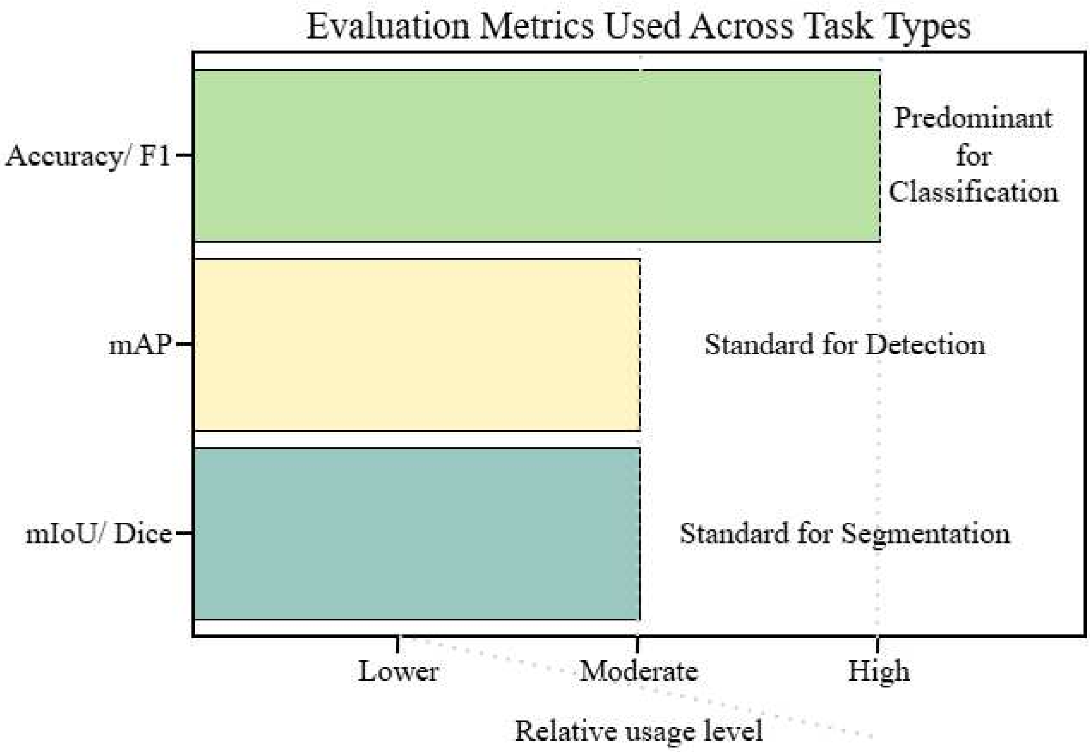

Figure 4: Metrics Usage

As illustrated in Table 5 and Figure 4, the absence of unified metrics limits the ability to benchmark models across different tasks.

4.4 Performance Comparison: Real vs Reported Performance

A critical finding across the reviewed literature is the gap between reported performance and real-world applicability. Most studies report very high accuracy (typically above 95%) when evaluated on laboratory datasets [50], [3], [45]. However, performance significantly decreases under real-world conditions due to environmental variability [53], [54]. Cross-dataset evaluation further reveals that models trained on one dataset often fail to generalize to others, indicating dataset-specific learning rather than true disease representation [57][57], [55].

Sr. No.	Evaluation Setting	Performance Trend	Interpretation
1	Laboratory	Very high accuracy	Overfitting risk
2	Field	Moderate accuracy	Realistic performance
3	Cross-dataset	Lower accuracy	Generalization issue

Table 6: Reported vs Real-World Performance

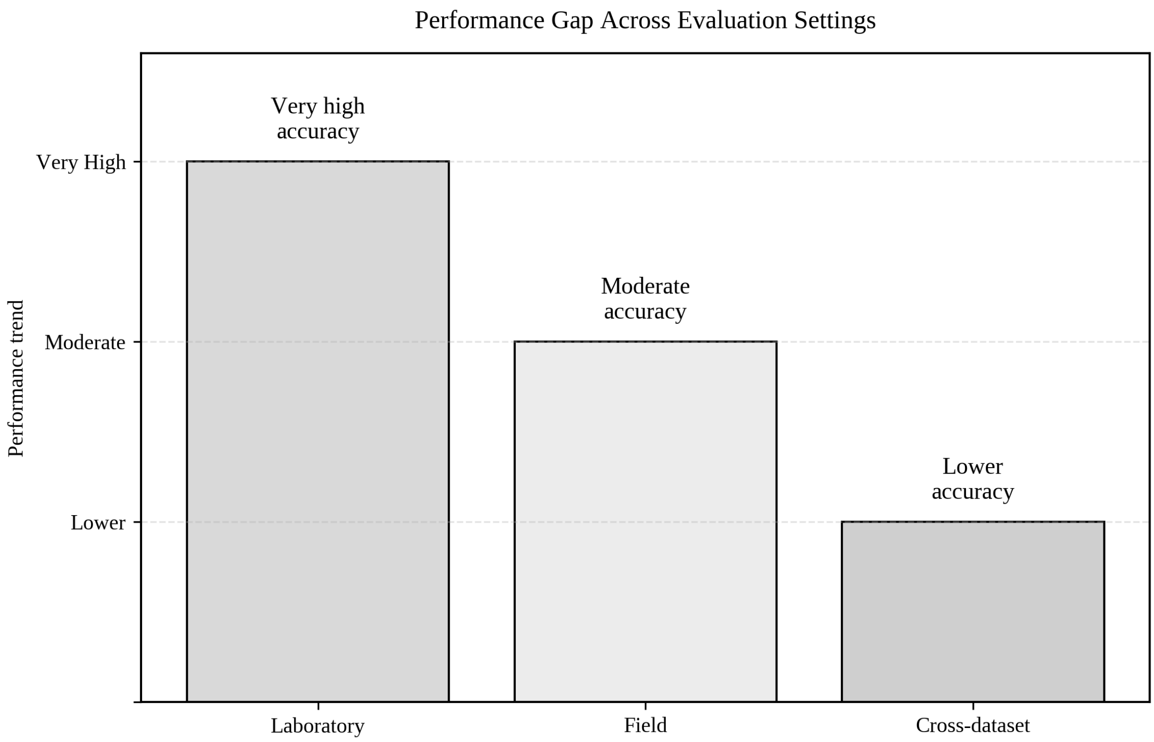

Figure 5: Performance Gap

The comparison in Table 6 and Figure 5 demonstrates that high benchmark accuracy does not guarantee real-world effectiveness.

4.5 Computational Efficiency

Computational efficiency is essential for deploying models in real-world agricultural environments, particularly in mobile and edge-based systems. Lightweight CNN models such as MobileNet and optimized architectures provide a balance between accuracy and efficiency, making them suitable for deployment [46], [39]. In contrast, Transformer and hybrid models, while more accurate, often require higher computational resources, limiting their practical applicability [47], [40].

Recent studies focus on improving efficiency through:

- Model compression
- Quantization
- Efficient attention mechanisms [36], [48]

Sr. No.	Model Type	Efficiency	Deployment Suitability
1	Lightweight CNN	High	Suitable for edge devices
2	Segmentation models	Moderate	Limited deployment
3	Hybrid/Transformer	Low	Resource-intensive

Table 7: Computational Efficiency Comparison

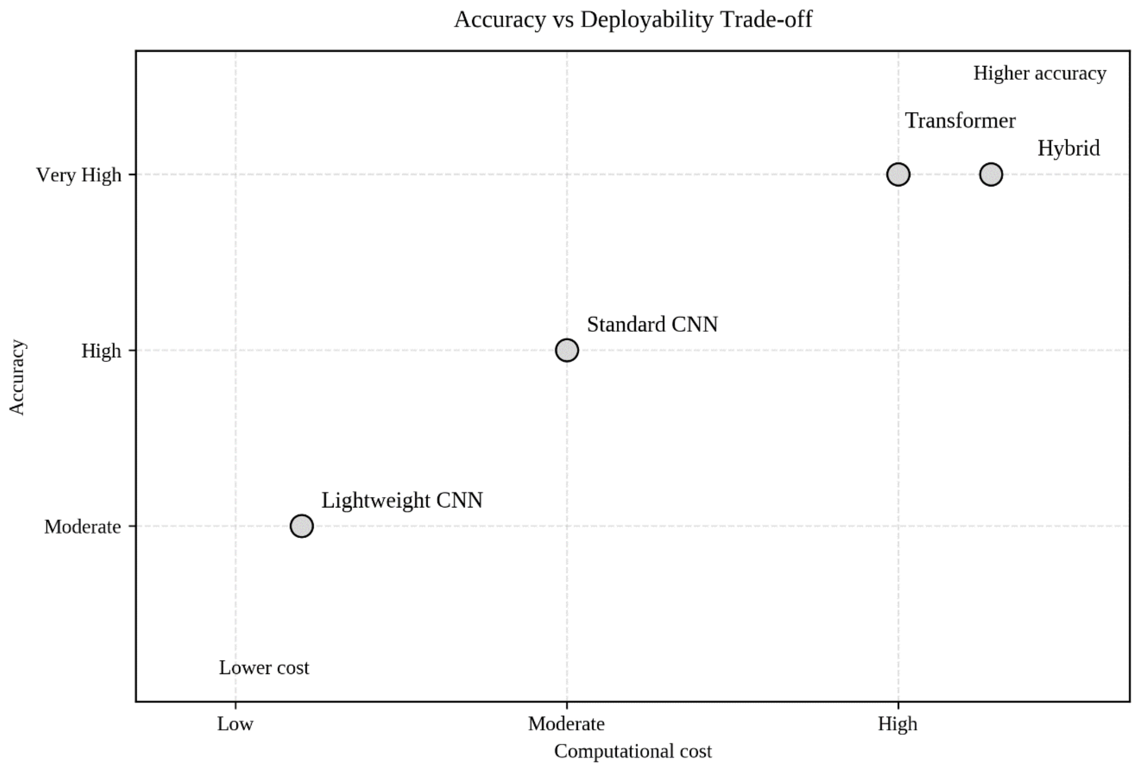

Figure 6: Trade-off Graph

Table 7 and Figure 6 highlights the trade-off between performance and deployment feasibility.

4.6 Synthesis of Findings

The comparative analysis across multiple dimensions reveals several important insights.

First, CNN-based classification models dominate the literature due to their simplicity and high accuracy; however, their performance is heavily dependent on controlled datasets, limiting real-world applicability. Second, Transformer and hybrid models improve robustness by incorporating global contextual information, but their adoption is constrained by computational complexity. Third, dataset selection plays a crucial role in model evaluation, with laboratory datasets leading to inflated performance and field datasets providing more realistic assessment. Fourth, the lack of standardized evaluation metrics makes cross-study comparison difficult and may lead to misleading conclusions. Finally, while lightweight models support deployment in resource-constrained environments, a trade-off between accuracy and efficiency remains unresolved.

Most importantly, very few studies explicitly address disease severity estimation, and it is often treated as a secondary task rather than a primary research objective. This highlights a critical research gap aligned with the objective of this review and underscores the need for integrated frameworks that combine detection, segmentation, and severity assessment under real-world conditions.

5. Benchmark Synthesis

This section presents a structured benchmark synthesis of the reviewed studies, enabling a comprehensive comparison of deep learning approaches for tomato disease detection and severity assessment. The synthesis focuses on model types, tasks, datasets, evaluation metrics, performance, and limitations.

5.1 Comprehensive Benchmark Comparison

Sr. No.	Model / Approach	Task	Dataset	Metric	Performance	Limitation
1	CNN (VGG, AlexNet)	Classification [64], [9]	PlantVillage	Accuracy	~95–97%	Overfitting, no localization

2	DenseNet / ResNet	Classification [43], [3]	PlantVillage	Accuracy	~98–99%	Poor real-world generalization
3	Lightweight CNN	Classification [46], [39]	PlantVillage + others	Accuracy	~98%	Limited feature depth
4	Transfer Learning CNN	Classification [44], [45]	Multi-source	Accuracy	~99%	Dataset dependency
5	Ensemble CNN	Classification [49], [8]	Large datasets	Accuracy/F1	~99%+	High complexity
6	YOLO-based models	Detection [39], [58]	Field + lab	mAP	High (~0.85–0.95)	Miss early-stage disease
7	Faster R-CNN	Detection [54], [65]	Field datasets	mAP	Improved vs baseline	Slower inference
8	U-Net	Segmentation [60], [61]	Leaf datasets	Dice/ mIoU	~95%+	High annotation cost
9	Attention U-Net	Segmentation [36], [48]	Tomato datasets	Dice	Improved accuracy	Computational overhead
10	Transformer models	Classification/ Detection [47], [36]	Mixed datasets	Accuracy/ mAP	~98–99%	High compute cost
11	Hybrid CNN + Transformer	Multi-task [40], [48]	Multi-source	Accuracy	~99%	Complex architecture
12	GAN-based models	Augmentation/ Classification [56], [57]	Small datasets	Accuracy	Improved performance	Synthetic bias risk
13	Semi-supervised models	Segmentation [55], [66]	Limited labeled data	mIoU	Improved generalization	Training complexity

Table 8: Benchmark Comparison of Deep Learning Models for Tomato Disease Detection

The comprehensive comparison in Table 8 indicates that CNN-based models dominate classification tasks with very high accuracy, particularly on controlled datasets such as PlantVillage [64], [43]. Detection models such as YOLO and Faster R-CNN improve practical applicability through localization but struggle with early-stage disease detection [39], [58]. Segmentation approaches provide precise pixel-level analysis, enabling downstream tasks such as severity estimation, although they require extensive annotated data [60], [61]. Hybrid and Transformer-based models achieve superior performance by combining local and global feature learning but introduce higher computational complexity [47], [40]. Overall, the table highlights a shift from simple classification toward more integrated and application-oriented approaches.

5.2 Best Models per Task

Sr. No.	Task	Best Model Type	Key Strength
1	Classification [43], [3]	DenseNet / EfficientNet	High accuracy
2	Detection [39], [54]	YOLO / Faster R-CNN	Real-time localization
3	Segmentation [60], [36]	U-Net variants	Pixel-level precision
4	Severity Estimation [49], [61]	Segmentation-based models	Quantitative analysis
5	Multi-task [40], [48]	Hybrid models	Integrated prediction

Table 9: Best Performing Models by Task

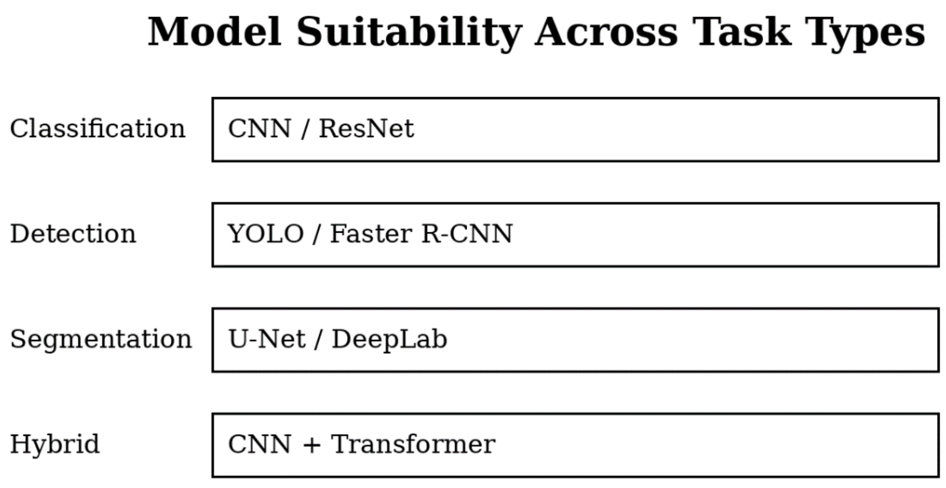

Figure 7: Different model architectures are optimal for different task types

Table 9 & Figure 7 identifies task-specific best-performing models, showing that CNN-based architectures remain dominant for classification due to their efficiency and accuracy [43], [3]. Detection models such as YOLO and Faster R-CNN are most suitable for real-time disease localization [39], [54], while U-Net-based architectures excel in segmentation tasks [60], [36].

Hybrid models demonstrate strong performance across multiple tasks, indicating their potential for integrated disease detection systems [40], [48]. Notably, severity estimation remains underrepresented, confirming a key research gap.

5.3 Best Dataset Analysis

Sr. No.	Dataset Type	Usage Frequency	Strength	Limitation
1	PlantVillage [50], [51]	Very High	Clean, high accuracy	Unrealistic
2	Field datasets [53], [54]	Moderate	Real-world applicability	Noise, complexity
3	Hybrid datasets [56], [57]	Increasing	Better generalization	Data inconsistency

Table 10: Dataset Benchmark Comparison

The dataset comparison in Table 10 reveals that most studies rely heavily on PlantVillage datasets due to their simplicity and availability, resulting in high reported accuracy [50], [51]. However, these datasets lack real-world variability, leading to poor generalization. Field datasets provide more realistic evaluation conditions but introduce challenges such as noise and annotation complexity [53], [54]. Hybrid datasets, which combine laboratory and field data, show improved robustness and generalization, suggesting a promising direction for future research [56], [57].

5.4 Real-World Ready Models

Sr. No.	Model Type	Feature	Suitability
1	Lightweight CNN [46], [39]	Low computation	High
2	YOLO-based models [39], [58]	Real-time detection	High
3	Hybrid models [40], [48]	High accuracy	Moderate
4	Transformer models [47], [36]	Robust features	Low

Table 11: Models Suitable for Real-World Deployment

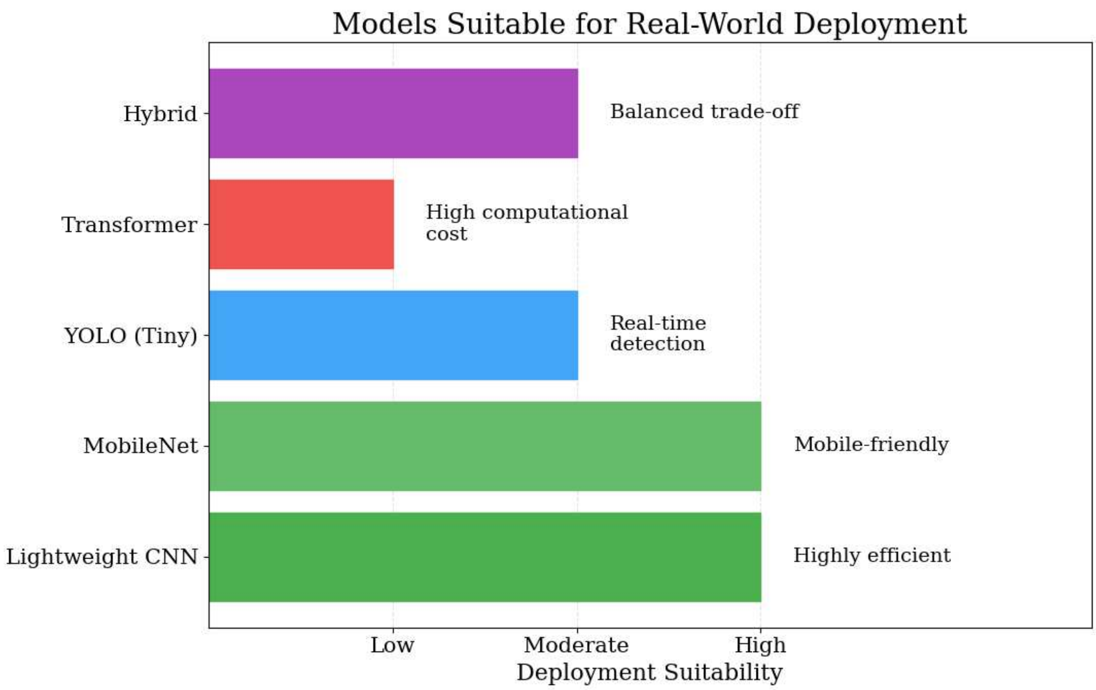

Figure 8: Models Suitable for Real-World Deployment

Table 11 & Figure 8, highlights & demonstrates that lightweight CNN and YOLO-based detection models are the most suitable for real-world deployment due to their efficiency and real-time capability [46], [39]. In contrast, Transformer and hybrid models, while highly accurate, are less suitable for deployment due to their computational requirements [47], [40]. This indicates that practical agricultural systems must balance performance with computational efficiency.

5.5 Lightweight vs Heavy Models

Sr. No.	Model Category	Accuracy	Computation	Deployment
1	Lightweight CNN [46], [39]	Moderate–High	Low	High
2	Standard CNN [43], [3], [44]	High	Moderate	Moderate
3	Transformer [47], [36]	Very High	High	Low
4	Hybrid [40], [48]	Very High	Very High	Moderate

Table 12: Computational Trade-off Analysis

Table 12 illustrates the trade-off between model complexity and deployment feasibility. Lightweight CNN models provide high efficiency and are well-suited for mobile and edge applications [46], [39]. Standard CNN models offer a balance between performance and computation [43], [3]. Transformer and hybrid models achieve the highest accuracy but require significantly higher computational resources, limiting their deployment in resource-constrained environments [47], [40]. This trade-off emphasizes the need for optimized models that maintain accuracy while reducing computational cost.

5.6 Benchmark Synthesis Insights

The benchmark synthesis reveals several critical observations:

- Most studies focus on classification tasks, with limited attention to detection and severity estimation (E1–E73).
- DenseNet and EfficientNet consistently achieve high accuracy but lack real-world robustness [43], [3].
- Detection and segmentation models provide better localization but require more annotated data [58], [60].
- Hybrid and Transformer-based models offer improved performance but increase computational complexity [47], [40].
- Lightweight models are essential for real-world deployment but may compromise accuracy [46], [39].
- Severity estimation remains underexplored and lacks standardized benchmarks [49], [61].

6. Research Gaps and Challenges

Despite significant advancements in deep learning-based tomato disease detection, several critical research gaps and challenges persist. These limitations, identified through the comparative and benchmark analysis of reviewed studies, hinder the practical deployment and scalability of existing approaches.

6.1 Over-Reliance on Laboratory Datasets

A major limitation across the literature is the heavy dependence on laboratory-based datasets, particularly PlantVillage-like datasets. These datasets provide controlled environments with uniform backgrounds and minimal noise, leading to inflated model performance [35], [16].

However, such conditions do not reflect real agricultural environments. Models trained on these datasets often fail when exposed to field conditions involving complex backgrounds, varying illumination, and occlusions [67], [68]. Furthermore, limited diversity in disease appearance and environmental variability restricts the model's ability to generalize across different geographical regions and crop conditions.

6.2 Lack of Severity Estimation

Although most studies focus on disease classification and detection, severity estimation remains largely underexplored. Only a limited number of works attempt to quantify the extent of infection, and even those approaches often rely indirectly on segmentation outputs rather than dedicated severity models [63], [69].

Additionally, there is a lack of:

- Standardized severity scales
- Benchmark datasets for severity assessment
- Evaluation metrics tailored for severity prediction

This gap is particularly critical, as severity estimation is essential for decision-making in precision agriculture, such as determining pesticide usage and crop management strategies.

6.3 Poor Real-World Generalization

Another significant challenge is the poor generalization of models in real-world scenarios. While many models report high accuracy on training datasets, their performance drops considerably when applied to unseen or cross-domain data [70], [71]. This issue arises due to:

- Domain shift between laboratory and field images
- Variability in environmental conditions
- Limited dataset diversity

Recent studies highlight that the models often learn dataset-specific features rather than disease-specific characteristics, resulting in reduced robustness and reliability in practical applications [72], [73].

6.4 Lack of Unified Benchmarking Framework

The absence of a standardized benchmarking framework is another critical limitation. Different studies use varying datasets, evaluation metrics, and experimental setups, making it difficult to perform fair comparisons across models [62], [74]. For instance:

- Classification studies report accuracy

- Detection studies use mAP
- Segmentation studies rely on mIoU or Dice

This inconsistency leads to fragmented research outcomes and limits the ability to identify truly superior models. A unified benchmark with standardized datasets and evaluation protocols is essential for advancing the field.

6.5 Limited Explainability and Interpretability

Most deep learning models used for tomato disease detection function as black-box systems, providing predictions without clear explanations. This lack of interpretability reduces trust and adoption among end-users, particularly farmers and agricultural practitioners [7], [75].

Although some recent studies incorporate explainable AI techniques, such as attention mechanisms and visualization methods, these approaches are still limited and not widely integrated into existing systems. Enhancing explainability is crucial for:

- Improving model transparency
- Supporting decision-making
- Increasing user acceptance

6.6 High Computational Cost

Advanced models, particularly Transformer-based and hybrid architectures, often involve high computational complexity, making them unsuitable for deployment in resource-constrained environments [66], [76]. Challenges include:

- Large model size
- High training and inference time
- Requirement of specialized hardware

Although lightweight models have been proposed, they often involve trade-offs between accuracy and efficiency. Achieving an optimal balance between performance and computational cost remains a key research challenge.

7. Future Research Directions

Based on the identified research gaps, several key directions can guide future advancements in tomato disease detection and severity assessment.

First, there is a strong need for multi-task learning frameworks that integrate disease detection, segmentation, and severity estimation within a unified model. Current approaches treat these tasks independently, limiting their effectiveness in real-world decision-making [71], [73].

Second, the development of large-scale field datasets is essential to improve model generalization. Existing datasets are heavily biased toward controlled environments, and future datasets must incorporate real-world variability such as lighting conditions, occlusions, and diverse crop stages [67], [68].

Third, research should focus on lightweight and deployable models suitable for edge devices, smartphones, and IoT-based agricultural systems. While high-performance models exist, their computational requirements restrict practical implementation [66], [76].

Fourth, the integration of explainable artificial intelligence (XAI) techniques is crucial to enhance model transparency and user trust. Interpretable models can support farmers and domain experts in understanding predictions and making informed decisions [7], [75].

Finally, future studies should emphasize cross-dataset validation and domain adaptation techniques to ensure robustness across different environmental conditions and datasets. This will reduce overfitting and improve real-world applicability [72].

8. Implications for Research Objective

This review directly supports the research objective of systematically analyzing deep learning methods for tomato disease detection and severity assessment. The findings highlight that while significant progress has been made in disease classification and detection, severity estimation remains insufficiently explored, despite its critical importance in precision agriculture.

Accurate severity assessment enables better decision-making regarding treatment strategies, resource allocation, and yield optimization. Furthermore, the analysis demonstrates the necessity of integrated multi-task frameworks that combine detection and severity estimation. Such approaches can bridge the gap between theoretical model performance and practical agricultural applications.

Therefore, this study establishes a clear foundation for future research focused on developing robust, interpretable, and deployable models that address both disease identification and severity quantification in real-world conditions.

9. Conclusion

This review presented a comprehensive and analytical evaluation of deep learning approaches for tomato disease detection, covering classification, detection, segmentation, and emerging multi-task methodologies.

The analysis revealed that CNN-based models dominate the field and achieve high accuracy, particularly on controlled datasets. However, these models suffer from limited generalization in real-world environments. Detection and segmentation methods improve localization and enable further analysis, yet they require more complex data and annotation. Hybrid and Transformer-based models demonstrate improved performance but introduce computational challenges. Several critical gaps were identified, including over-reliance on laboratory datasets, lack of standardized benchmarks, limited focus on severity estimation, and challenges in model deployment and interpretability.

Overall, this study emphasizes the importance of developing integrated, robust, and efficient deep learning frameworks that move beyond classification toward comprehensive disease analysis, including severity estimation. Addressing these challenges is essential for advancing precision agriculture and ensuring sustainable crop management.

10. Statements and Declarations

Competing Interests: The authors declare that they have no known financial or non-financial competing interests that could have appeared to influence the work reported in this paper.

Funding: The authors received no specific funding for this work.

Employment: The authors are not employed by any organization that could gain or lose financially through the publication of this manuscript.

Financial Interests: The authors declare that they have no financial interests, including stocks, patents, or consultancies, that could be affected by the publication of this work.

Non-Financial Interests: The authors declare that they have no relevant non-financial interests, including personal relationships or professional affiliations, that could influence the work reported in this manuscript.

Data Availability: No new data were created or analyzed in this study as this is a review paper.

Author Contributions: Shubham Sandal conducted the literature review, analysis, and manuscript preparation. Dr. Sukanta Ghosh provided supervision, guidance, and critical review of the manuscript.

11. References

[1] N. K. Trivedi *et al.*, “Early Detection and Classification of Tomato Leaf Disease Using High-Performance Deep Neural Network,” *Sensors*, vol. 21, no. 23, p. 7987, Nov. 2021, doi: 10.3390/s21237987.

[2] M. Nawaz *et al.*, “A robust deep learning approach for tomato plant leaf disease localization and classification,” *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21498-5.

[3] J. Liu and X. Wang, “Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network,” *Front. Plant Sci.*, vol. 11, Jun. 2020, doi: 10.3389/fpls.2020.00898.

[4] Y. Liu *et al.*, “High-Precision Tomato Disease Detection Using NanoSegmenter Based on Transformer and Lightweighting,” *Plants*, vol. 12, no. 13, p. 2559, Jul. 2023, doi: 10.3390/plants12132559.

[5] L. Li, S. Zhang, and B. Wang, “Plant Disease Detection and Classification by Deep Learning—A Review,” *IEEE Access*, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[6] N. Khasawneh, E. Faouri, and M. Fraiwan, “Automatic Detection of Tomato Diseases Using Deep Transfer Learning,” *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep. 2022, doi: 10.3390/app12178467.

[7] A. Yu, Y. Xiong, Z. Lv, P. Wang, J. She, and L. Wei, “New Method for Tomato Disease Detection Based on Image Segmentation and Cycle-GAN Enhancement,” *Sensors*, vol. 24, no. 20, p. 6692, Oct. 2024, doi: 10.3390/s24206692.

[8] M. Jelali, “Deep learning networks-based tomato disease and pest detection: a first review of research studies using real field datasets,” *Front. Plant Sci.*, vol. 15, Oct. 2024, doi: 10.3389/fpls.2024.1493322.

[9] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, “Recent advances in image processing techniques for automated leaf pest and disease recognition – A review,” *Information Processing in Agriculture*, vol. 8, no. 1, pp. 27–51, Mar. 2021, doi: 10.1016/j.inpa.2020.04.004.

[10] W. Hu, W. Hong, H. Wang, M. Liu, and S. Liu, “A Study on Tomato Disease and Pest Detection Method,” *Applied Sciences (Switzerland)*, vol. 13, no. 18, Sep. 2023, doi: 10.3390/app131810063.

[11] D. Karimanzira, “Context-Aware Tomato Leaf Disease Detection Using Deep Learning in an Operational Framework,” *Electronics (Switzerland)*, vol. 14, no. 4, Feb. 2025, doi: 10.3390/electronics14040661.

[12] J. U. M. Akbar, S. F. Kamarulzaman, A. J. M. Muzahid, M. A. Rahman, and M. Uddin, “A Comprehensive Review on Deep Learning Assisted Computer Vision Techniques for Smart Greenhouse Agriculture,” *IEEE Access*, vol. 12, pp. 4485–4522, 2024, doi: 10.1109/ACCESS.2024.3349418.

[13] M. Shoaib *et al.*, “An advanced deep learning models-based plant disease detection: A review of recent research,” *Front. Plant Sci.*, vol. 14, Mar. 2023, doi: 10.3389/fpls.2023.1158933.

[14] K. Joshi *et al.*, “Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation,” *Curr. Plant Biol.*, vol. 41, Mar. 2025, doi: 10.1016/j.cpb.2025.100437.

[15] Z. Ullah, N. Alsubaie, M. Jamjoom, S. H. Alajmani, and F. Saleem, “EffiMob-Net: A Deep Learning-Based Hybrid Model for Detection and Identification of Tomato Diseases Using Leaf Images,” *Agriculture (Switzerland)*, vol. 13, no. 3, Mar. 2023, doi: 10.3390/agriculture13030737.

[16] H. Dong *et al.*, “LPDiag: LLM-Enhanced Multimodal Prototype Learning Framework for Intelligent Tomato Leaf Disease Diagnosis,” *Agriculture*, vol. 16, no. 4, p. 419, Feb. 2026, doi: 10.3390/agriculture16040419.

[17] X. Wang and J. Liu, “An efficient deep learning model for tomato disease detection,” *Plant Methods*, vol. 20, no. 1, Dec. 2024, doi: 10.1186/s13007-024-01188-1.

[18] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, “A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition,” *Sensors (Switzerland)*, vol. 17, no. 9, Sep. 2017, doi: 10.3390/s17092022.

[19] L. Tan, J. Lu, and H. Jiang, “Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods,” *AgriEngineering*, vol. 3, no. 3, pp. 542–558, Sep. 2021, doi: 10.3390/agriengineering3030035.

[20] M. Shoaib *et al.*, “Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease,” *Front. Plant Sci.*, vol. 13, Oct. 2022, doi: 10.3389/fpls.2022.1031748.

[21] Z. Zhai, J. F. Martínez Ortega, N. L. Martínez, and J. Rodríguez-Molina, “A mission planning approach for precision farming systems based on multi-objective optimization,” *Sensors (Switzerland)*, vol. 18, no. 6, Jun. 2018, doi: 10.3390/s18061795.

[22] D. V. Leite, A. V. de Brito, G. G. Faccioli, and G. Haddad Souza Vieira, “Deep Learning Models for Detection and Severity Assessment of Cercospora Leaf Spot (Cercospora capsici) in Chili Peppers Under Natural Conditions,” *Plants*, vol. 14, no. 13, Jul. 2025, doi: 10.3390/plants14132011.

[23] M. Rezaei, D. Diepeveen, H. Laga, and F. Sohel, “Automatic pixel-level annotation for plant disease severity estimation,” *Comput. Electron. Agric.*, vol. 241, p. 111316, Feb. 2026, doi: 10.1016/j.compag.2025.111316.

[24] M. V. Sanida, T. Sanida, A. Sideris, and M. Dasygenis, “An Efficient Hybrid CNN Classification Model for Tomato Crop Disease,” *Technologies (Basel).*, vol. 11, no. 1, Feb. 2023, doi: 10.3390/technologies11010010.

[25] Y. Zhang, C. Song, and D. Zhang, “Deep Learning-Based Object Detection Improvement for Tomato Disease,” *IEEE Access*, vol. 8, pp. 56607–56614, 2020, doi: 10.1109/ACCESS.2020.2982456.

[26] S. Panno *et al.*, “A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin,” Nov. 01, 2021, *MDPI*. doi: 10.3390/agronomy11112188.

[27] Y. Lu and S. Young, “A survey of public datasets for computer vision tasks in precision agriculture,” Nov. 01, 2020, *Elsevier B.V.* doi: 10.1016/j.compag.2020.105760.

[28] A. Das, F. Pathan, J. R. Jim, M. M. Kabir, and M. F. Mridha, “Deep learning-based classification, detection, and segmentation of tomato leaf diseases: A state-of-the-art review,” Jun. 01, 2025, *KeAi Communications Co.* doi: 10.1016/j.aiia.2025.02.006.

[29] Y. Zhong, Z. Teng, and M. Tong, “LightMixer: A novel lightweight convolutional neural network for tomato disease detection,” *Front. Plant Sci.*, vol. 14, 2023, doi: 10.3389/fpls.2023.1166296.

[30] M. Xu, J. E. Park, J. Lee, J. Yang, and S. Yoon, “Plant disease recognition datasets in the age of deep learning: challenges and opportunities,” *Front. Plant Sci.*, vol. 15, 2024, doi: 10.3389/fpls.2024.1452551.

[31] A. Ahmad, D. Saraswat, and A. El Gamal, “A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools,” Feb. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.atech.2022.100083.

[32] M. S. Alzahrani and F. W. Alsaade, “Transform and Deep Learning Algorithms for the Early Detection and Recognition of Tomato Leaf Disease,” *Agronomy*, vol. 13, no. 5, May 2023, doi: 10.3390/agronomy13051184.

[33] I. Pacal *et al.*, “A systematic review of deep learning techniques for plant diseases,” *Artif. Intell. Rev.*, vol. 57, no. 11, Nov. 2024, doi: 10.1007/s10462-024-10944-7.

[34] E. Ozbilge, M. K. Ulukok, O. Toygar, and E. Ozbilge, “Tomato Disease Recognition Using a Compact Convolutional Neural Network,” *IEEE Access*, vol. 10, pp. 77213–77224, 2022, doi: 10.1109/ACCESS.2022.3192428.

[35] A. K. Shahade and P. V. Deshmukh, “Edge-enhanced dual branch CNN with adaptive attention for robust apple leaf disease detection,” *BMC Plant Biol.*, vol. 26, no. 1, p. 39, Dec. 2025, doi: 10.1186/s12870-025-07851-6.

[36] A. Batool, J. Kim, and Y.-C. Byun, "A compact deep learning approach integrating depthwise convolutions and spatial attention for plant disease classification," *Plant Methods*, vol. 21, no. 1, p. 48, Apr. 2025, doi: 10.1186/s13007-025-01325-4.

[37] P. Pai, N. S. J., S. T., A. S., and N. Gondi, "Intelligent plant disease detection using twin attention optimal convolutional neural network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 15, no. 1, p. 756, Feb. 2026, doi: 10.11591/ijai.v15.i1.pp756-765.

[38] M. Shafay *et al.*, "Recent advances in plant disease detection: challenges and opportunities," *Plant Methods*, vol. 21, no. 1, p. 140, Oct. 2025, doi: 10.1186/s13007-025-01450-0.

[39] Y. Zhang, C. Song, and D. Zhang, "Deep Learning-Based Object Detection Improvement for Tomato Disease," *IEEE Access*, vol. 8, pp. 56607–56614, 2020, doi: 10.1109/ACCESS.2020.2982456.

[40] S. Abinaya, K. U. Kumar, and A. S. Alphonse, "Cascading Autoencoder With Attention Residual U-Net for Multi-Class Plant Leaf Disease Segmentation and Classification," *IEEE Access*, vol. 11, pp. 98153–98170, 2023, doi: 10.1109/ACCESS.2023.3312718.

[41] Y. Dan *et al.*, "DKP-ADS: Domain knowledge prompt combined with multi-task learning for assessment of foliar disease severity in staple crops," *Crop J.*, vol. 13, no. 6, pp. 1939–1954, Dec. 2025, doi: 10.1016/j.cj.2025.09.017.

[42] M. Khalid and MD. A. Talukder, "A Hybrid Deep Multistacking Integrated Model for Plant Disease Detection," *IEEE Access*, vol. 13, pp. 116037–116053, 2025, doi: 10.1109/ACCESS.2025.3583793.

[43] R. Sujatha, J. M. Chatterjee, N. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocess. Microsyst.*, vol. 80, p. 103615, Feb. 2021, doi: 10.1016/j.micpro.2020.103615.

[44] M. E. H. Chowdhury *et al.*, "Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques," *AgriEngineering*, vol. 3, no. 2, pp. 294–312, May 2021, doi: 10.3390/agriengineering3020020.

[45] E. Moupojou *et al.*, "FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning," *IEEE Access*, vol. 11, pp. 35398–35410, 2023, doi: 10.1109/ACCESS.2023.3263042.

[46] M. H. Saleem, J. Potgieter, and K. M. Arif, "Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments," *Precis. Agric.*, vol. 22, no. 6, pp. 2053–2091, Dec. 2021, doi: 10.1007/s11119-021-09806-x.

[47] S. Hernández and J. L. López, "Uncertainty quantification for plant disease detection using Bayesian deep learning," *Appl. Soft Comput.*, vol. 96, p. 106597, Nov. 2020, doi: 10.1016/j.asoc.2020.106597.

[48] M. Z. Muntaqim, H. M. Kafi, and T. A. Smrity, “Privacy-aware plant disease detection: federated learning with homomorphic encryption on image data,” *Vis. Comput.*, vol. 42, no. 1, p. 18, Jan. 2026, doi: 10.1007/s00371-025-04261-5.

[49] D. Argüeso *et al.*, “Few-Shot Learning approach for plant disease classification using images taken in the field,” *Comput. Electron. Agric.*, vol. 175, p. 105542, Aug. 2020, doi: 10.1016/j.compag.2020.105542.

[50] K. R., H. M., S. Anand, P. Mathikshara, A. Johnson, and M. R., “Attention embedded residual CNN for disease detection in tomato leaves,” *Appl. Soft Comput.*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/j.asoc.2019.105933.

[51] P. Sharma, Y. P. S. Berwal, and W. Ghai, “Performance analysis of deep learning CNN models for disease detection in plants using image segmentation,” *Information Processing in Agriculture*, vol. 7, no. 4, pp. 566–574, Dec. 2020, doi: 10.1016/j.inpa.2019.11.001.

[52] M. Loey, A. ElSawy, and M. Afify, “Deep Learning in Plant Diseases Detection for Agricultural Crops,” *International Journal of Service Science, Management, Engineering, and Technology*, vol. 11, no. 2, pp. 41–58, Apr. 2020, doi: 10.4018/IJSSMET.2020040103.

[53] V. Tiwari, R. C. Joshi, and M. K. Dutta, “Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images,” *Ecol. Inform.*, vol. 63, p. 101289, Jul. 2021, doi: 10.1016/j.ecoinf.2021.101289.

[54] J. H. Sinamenye, A. Chatterjee, and R. Shrestha, “Potato plant disease detection: leveraging hybrid deep learning models,” *BMC Plant Biol.*, vol. 25, no. 1, p. 647, May 2025, doi: 10.1186/s12870-025-06679-4.

[55] A. Şener and B. Ergen, “Advanced CNN Approach for Segmentation of Diseased Areas in Plant Images,” *Journal of Crop Health*, vol. 76, no. 6, pp. 1569–1583, Dec. 2024, doi: 10.1007/s10343-024-01054-z.

[56] A. Pal and V. Kumar, “AgriDet: Plant Leaf Disease severity classification using agriculture detection framework,” *Eng. Appl. Artif. Intell.*, vol. 119, p. 105754, Mar. 2023, doi: 10.1016/j.engappai.2022.105754.

[57] A. Bilal, J. A. Khan, A. Alzahrani, K. Almohammadi, M. Alamri, and X. Liu, “Fuzzy deep learning architecture for cucumber plant disease detection and classification,” *J. Big Data*, vol. 12, no. 1, p. 117, May 2025, doi: 10.1186/s40537-025-01156-z.

[58] U. Barman *et al.*, “ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture,” *Agronomy*, vol. 14, no. 2, p. 327, Feb. 2024, doi: 10.3390/agronomy14020327.

[59] T. Shi *et al.*, “Recent advances in plant disease severity assessment using convolutional neural networks,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-29230-7.

[60] H. Bakır, “Evaluating the impact of tuned pre-trained architectures’ feature maps on deep learning model performance for tomato disease detection,” *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 18147–18168, Nov. 2023, doi: 10.1007/s11042-023-17503-2.

[61] M. A. W. Nazri and T. R. Razak, “Towards deployable and explainable deep learning models for paddy leaf disease classification in R: A comparative study of CNN architectures with SHAP and LIME,” *Expert Syst. Appl.*, vol. 300, p. 130337, Mar. 2026, doi: 10.1016/j.eswa.2025.130337.

[62] B. Liu, “Automatic Identification of Tomato Disease Based on Deep Learning,” *Pakistan Journal of Life and Social Sciences (PJLSS)*, vol. 22, no. 2, 2024, doi: 10.57239/PJLSS-2024-22.2.00738.

[63] A. S. Rathor *et al.*, “Deep convolutional network for plant monitoring in precision agriculture,” *Sustainable Computing: Informatics and Systems*, vol. 50, p. 101302, Jun. 2026, doi: 10.1016/j.suscom.2026.101302.

[64] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, “Using deep transfer learning for image-based plant disease identification,” *Comput. Electron. Agric.*, vol. 173, p. 105393, Jun. 2020, doi: 10.1016/j.compag.2020.105393.

[65] H. Qiu, J. Yang, J. Jiang, and W. Zhang, “Mob-psp: modified MobileNet-V2 network for real-time detection of tomato diseases,” *J. Real. Time. Image Process.*, vol. 21, no. 5, p. 181, Oct. 2024, doi: 10.1007/s11554-024-01561-2.

[66] B. Ranjitha and A. K. Sampath, “Deep learning-based hybrid optimisation for multiclass plant disease detection using leaf images in a distributed environment,” *International Journal of Intelligent Information and Database Systems*, vol. 18, no. 1, pp. 48–82, 2026, doi: 10.1504/IJIIDS.2026.150437.

[67] D. Wadhwa and K. Malik, “Deep learning generalized hybrid models for multi-species crop disease classification with explainable insights,” *Eng. Appl. Artif. Intell.*, vol. 164, p. 113317, Jan. 2026, doi: 10.1016/j.engappai.2025.113317.

[68] S. Venkatraman, P. K. Sathya Venkatesh, P. Vairamoorthy, A. Anandakumar, and A. Ashok S., “SPROUT: Symptom-centric Prototypical Representation Optimization and Uncertainty-aware Tuning for few-shot precision agriculture,” *Neurocomputing*, vol. 667, p. 132137, Feb. 2026, doi: 10.1016/j.neucom.2025.132137.

[69] J. Dhannya and Dr. S. Veni, “Deep Neural Networks for Plant Disease Detection: Insights from VGG16, U-Net, and ResNet,” *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.*, vol. 17, no. 1, pp. 34–46, Mar. 2026, doi: 10.58346/JOWUA.2026.I1.003.

[70] Y. M. Gajmal, A. M. Jagtap, K. D. Kale, J. S. Gawade, and P. More, “A Blockchain-Based Hybrid Hunger Game Search Archimedes Optimization Enabled Deep Learning for Multiclass Plant Disease Detection Using Leaf Images,” *Int. J. Image Graph.*, vol. 26, no. 03, May 2026, doi: 10.1142/S021946782650018X.

[71] J. Rahaman, P. Paul, A. Chowdhury, and M. Quamruzzaman, “A customized MobileNetV3Large-based deep learning framework for plant disease detection,”

Discover Artificial Intelligence, vol. 6, no. 1, p. 110, Jan. 2026, doi: 10.1007/s44163-025-00733-8.

[72] A. Al Mamun, M. Zhang, D. Ahmedt-Aristizabal, Z. Hayder, and M. Awrangjeb, “ConMamba: Contrastive vision Mamba for plant disease detection,” *Pattern Recognit.*, vol. 176, p. 113177, Aug. 2026, doi: 10.1016/j.patcog.2026.113177.

[73] C. R. Medentzidis, D. Kapetas, and E. M. Pechlivani, “Fungal Disease Detection on Vegetable Crops Using RGB Imaging,” *Agriculture*, vol. 16, no. 5, p. 541, Feb. 2026, doi: 10.3390/agriculture16050541.

[74] D. Zaibi, A. Alkameli, and R. Ksantini, “An Enhanced Dynamic Autoencoder Based on Gaussian Mixture Model for Plant Disease Detection,” *Comput. Intell.*, vol. 42, no. 1, Feb. 2026, doi: 10.1111/coin.70164.

[75] V. Kumari and V. Bansal, “A systematic review of plant leaf disease detection and classification using machine learning and deep learning techniques,” *Int. J. Data Sci. Anal.*, vol. 22, no. 1, p. 22, Dec. 2026, doi: 10.1007/s41060-025-01003-7.

[76] C.-F. Long, S. Li, H.-J. Ao, Y.-J. Deng, T. Hu, and Z.-H. Li, “A ConvNeXt–LiteMamba Dual-Branch Network for Detection of Rice Blast Disease via Hyperspectral Imaging,” *Agronomy*, vol. 16, no. 5, p. 500, Feb. 2026, doi: 10.3390/agronomy16050500.