Deep Learning-Based Automated Tomato Detection and Ripeness Classification Using Convolutional Neural Networks

1st **Dr. V T Ram Pavan Kumar M**
Associate Professor
Department of Computer Science
Kakaraparti Bhavanarayana College
mrpphd2018@gmail.com

2nd **Dr. R. Kiran Kumar**
Associate Professor
Department of Computer Science
Krishna University
Machilipatnam, Andhra Pradesh
Kirankreddi@gmail.com

3rd **P. L. Ramesh**
Assistant Professor
Department of Computer Science
Kakaraparti Bhavanarayana College
Rameshplus@yahoo.com

4th **Renuka Kummari**
Assistant Professor
Department of Computer Science and Engineering
CMR Institute of Technology
Hyderabad, Telangana, India
renukam24@gmail.com

5th **Valluri Lakshmi Sravani**
BTech Student
Department of CSE-AIML
Gitam University
Bengaluru
sravani.pardhasaradhi@gmail.com

6th **Sriharsha Vikruthi***
Assistant Professor
Department of Computer Science and Engineering
B V Raju Institute of Technology
Narsapur, Medak, Telangana, India 502313
sriharsha.vikruthi@gmail.com

***Abstract*—Tomato detection and ripeness classification are essential tasks in precision agriculture, as they support automated harvesting, quality grading, and post-harvest management. Traditional manual inspection methods are labor-intensive, subjective, and often produce inconsistent results under varying environmental conditions. To address these limitations, this paper proposes a deep learning-based framework for automated tomato detection and ripeness classification using Convolutional Neural Networks (CNNs). The proposed system integrates image preprocessing, data augmentation, tomato detection, deep feature extraction, and transfer learning-based classification to accurately identify tomatoes and categorize them into four ripeness stages: Unripe, Breaker, Semi-Ripe, and Fully Ripe. A pretrained ResNet-50 architecture is employed to extract discriminative features related to color, texture, and shape, enabling robust classification under complex agricultural environments. The performance of the proposed framework is evaluated using standard metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Experimental results demonstrate that the proposed approach achieves an accuracy of 96.84%, precision of 95.92%, recall of 95.47%, F1-score of 95.69%, and ROC-AUC of 98.12%, outperforming several existing methods. The developed framework provides an efficient and reliable solution for intelligent harvesting and automated grading systems, thereby reducing manual effort and enhancing productivity in smart farming applications.**

***Index Terms*—Tomato Detection, Ripeness Classification, Convolutional Neural Networks (CNN), Deep Learning, Transfer Learning, ResNet-50, Computer Vision, Precision Agriculture, Smart Farming, Automated Harvesting.**

I. INTRODUCTION

Tomatoes (*Solanum lycopersicum* L.) are one of the most widely cultivated vegetable crops worldwide and play a vital role in human nutrition and the agricultural economy. Their quality and market value largely depend on the stage of ripeness at which they are harvested. Accurate identification of tomato maturity is essential for ensuring optimal taste, nutritional quality, shelf life, and suitability for transportation and processing. Consequently, efficient tomato detection and ripeness assessment have become important components of modern precision agriculture.

Conventionally, tomato ripeness is determined through manual visual inspection based on external characteristics such as color and appearance. Although this method is simple, it is highly dependent on human expertise and is often subjective, time-consuming, labor-intensive, and inconsistent. In large-scale farming environments, manual assessment becomes impractical due to the increasing demand for high productivity and the shortage of skilled labor. Moreover, environmental factors such as varying illumination conditions, occlusions caused by leaves and branches, and complex backgrounds further complicate the ripeness evaluation process.

Recent advancements in artificial intelligence, computer vision, and deep learning have provided promising solutions to these challenges. In particular, Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image recognition tasks due to their ability to automatically learn discriminative features from raw image data. The application of CNNs in agriculture has enabled automated fruit detection, disease diagnosis, yield estimation, and maturity classification with improved accuracy and reliability. These intelligent systems minimize human intervention and support efficient decision-making in agricultural operations.

Automated tomato detection and ripeness classification can significantly improve harvesting efficiency, fruit grading, and post-harvest management. By accurately identifying maturity stages, farmers can harvest tomatoes at the optimal time, thereby reducing post-harvest losses and improving product quality. Furthermore, integrating deep learning techniques into agricultural systems contributes to the development of smart farming practices that enhance productivity while reducing operational costs.

In this work, a deep learning-based framework for automated tomato detection and ripeness classification using Convolutional Neural Networks is proposed. The framework integrates image preprocessing, data augmentation, tomato detection, transfer learning-based feature extraction, and classification to accurately categorize tomatoes into four ripeness stages: Unripe, Breaker, Semi-Ripe, and Fully Ripe. A pre-trained ResNet-50 architecture is employed to extract deep visual features and perform classification under varying environmental conditions. The proposed approach aims to provide an efficient, robust, and reliable solution for intelligent harvesting and quality assessment applications in precision agriculture.

The main objectives of this study are to develop an automated system capable of accurately detecting tomatoes from images, classifying them according to their ripeness stages, and evaluating the effectiveness of the proposed framework using standard performance metrics. The developed system has the potential to support automated harvesting robots, fruit grading systems, and decision-support tools for sustainable smart farming applications.

II. Literature Survey

Out in fields where light shifts and leaves tangle, spotting ripe fruit once slowed workers. Today's systems learn patterns through layers of digital thought, shaped by tons of labeled images. Where older methods failed under shadow or dirt, newer ones adjust using spatial awareness built from training runs. One approach leans on grid-like filters sweeping across scenes, another pulls in long-range context like a reader scanning paragraphs. Even when plants crowd together, some setups track each round shape without losing count. Speed matters just as much as accuracy, especially during harvest rushes. Designs inspired by fast object detectors now run on lightweight gear mounted to tractors or drones. Not every system works the same way - some build maps of color patches first, then decide which blobs are actual tomatoes. Through repeated trials, they get less confused by red weeds or wet soil glare. Performance holds up better now, even at odd angles or half-hidden behind stems.

From ResNet-152 came a creation by Mutahar's group - a way to judge tomato ripeness. It does not rely on preset features but discovers intricate visual details independently. What stands out? The method cleanly divided tomatoes into one of three stages: edible, unripe, or rotten. Thanks to internal skip paths in the design, results showed a solid jump in precision when evaluated.

Out of step with usual methods, Idakwo and colleagues [2] tried capsule networks since standard CNNs tend to blur spatial feature positions. With that change came better clarity - precision reached 99.56%. Because of it, judging fruit maturity grew steadier, particularly during automated grading. Without fanfare, the approach did more than earlier versions ever managed.

Lately, folks have started paying more attention to transformer-based models. Rather than sticking to regular network types, Nahak and colleagues [3] linked Vision Transformers with CNNs to judge how ripe fruits are - clearly telling apart under-ripened, mid-stage, and fully mature ones, hitting solid outcomes. Much like that, Khan et al. [11], [12], [13], [15] wove convolutional blocks together with transformer reasoning to handle tough visual setups: obscured sightlines, patchy lighting, messy backdrops. During trials focused on grouping and defining shapes in pictures, these mixed architectures outperformed earlier CNN-only versions, mainly because they notice broad arrangements within image sets. That advantage showed up when they picked up on far-off details many approaches tend to overlook.

One way people detect tomatoes quickly today involves systems built like YOLO. Not sticking to basic YOLOv8, Yang's crew slipped in clever attention modules along with a tougher feature merger - accuracy climbed to 95.8 percent. Following that path, Mali's team used YOLOv5 to locate fruit and guess ripeness, turning in strong real-time performance. Another angle showed up in paper [9], where scientists wove CBAM blocks into YOLOv5, helping it grab small or clustered tomatoes more reliably, lifting average precision to 88.1.

Fresh work has carried YOLO further than before. Rather than reuse past frameworks, Akbar and Kamarulzaman [10] wove in a Swin Transformer piece into YOLOv8s - this nudged up mAP@0.5 marks. In step with that thinking, Yan et al. [16] introduced YOLOv11-SLBA, adjusting feature merging methods so ripeness detection works better amid clutter. Elsewhere, Nahiduzzaman et al. [17] ran YOLOv8 on a Raspberry Pi, catching signs of tomato readiness fast, showing compact gear isn't too weak for farm smarts. At the same time, Liu [18] sharpened YOLOv10's vision using attention layers, boosting its mAP50-95 outcomes noticeably.

Not stopping at detection alone, scientists began classifying plant conditions while measuring progression stages. Built by Kebir's group [4], a model based on convolutional neural networks detected tomato ailments with high precision and dependable recall rates. Instead of texture or shape, Ayunda and others [8] relied mostly on skin tone through similar network designs, reaching almost 90 percent accuracy in assessing fruit ripeness. Going further, results climbed toward 97 percent in research by Putri and Rozi [19], again centered on surface color via CNN techniques. Even when surroundings turned chaotic, Mousse and collaborators [14] managed to distinguish five separate growth levels using deep learning approaches.

Chen and team [20] explored detecting tomato clusters in single greenhouses with a system named AFBF-YOLO; it spots cherry tomato groups, then assesses maturity by sharpening focus via enhanced attention blocks mixed with combined feature inputs for better location precision. Though intricate, the approach leans into subtle distinctions instead of wide generalizations - built on stacking data layers carefully, tuning outputs without falling back on one-size-fits-all templates across different settings.

Peek inside recent research and you see transformer models paired with YOLO frameworks doing solid detection work, especially in cluttered or rapidly shifting environments. Yet nearly all such approaches lean hard on strong processors while needing precise tuning steps. Top-tier tools might win benchmarks but often ignore basic realities - say, rural farms running outdated hardware. The gap? Slimmed-down CNN structures that still perform reliably without guzzling energy. Out there where machines struggle, getting by efficiently counts as much as hitting exact marks.

A fresh approach to spotting tomatoes uses deep learning, where Convolutional Neural Networks take charge of telling ripe from unripe without help. Still, speed meets simplicity on purpose, shaped for tools farmers can actually use. While fancier systems are out there, lean structure wins when bringing smarts into fields. High precision holds firm even though the gear needed stays light. Still, the design skips extra levels along with complex setups. Because of this, combining elements in the field feels doable when things are normal.

III. Proposed Methodology

The proposed methodology presents a deep learning-based framework for automated tomato detection and ripeness classification using Convolutional Neural Networks (CNNs). The framework integrates multiple stages, including dataset collection, preprocessing, data augmentation, tomato detection, deep feature extraction, ripeness classification, model optimization, and performance evaluation. The overall objective of the proposed system is to accurately detect tomatoes from images and classify them into different maturity stages, thereby supporting automated harvesting and quality assessment in smart agriculture.

A. Dataset Collection

The first stage of the proposed framework involves collecting tomato images from publicly available agricultural datasets, such as Kaggle repositories and image databases related to fruit maturity analysis. The dataset contains tomato samples belonging to four ripeness categories: Unripe (Green), Breaker (Turning), Semi-Ripe (Orange), and Fully Ripe (Red).

To improve the generalization capability of the model, images captured under different environmental conditions are included. These variations include changes in illumination, viewing angles, partial occlusions caused by leaves and branches, and complex backgrounds commonly observed in agricultural fields. A diverse dataset enables the model to learn representative features and improves its performance when deployed in real-world scenarios.

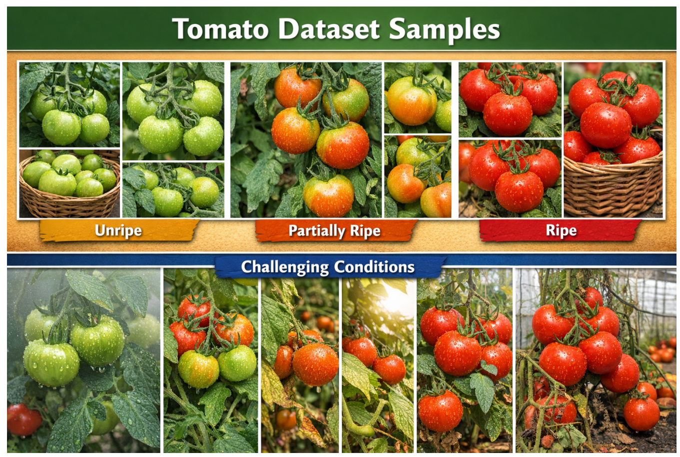

Fig. 1: Dataset

B. Image Preprocessing

Before training the deep learning model, the collected images undergo several preprocessing operations to standardize the input data and improve learning efficiency. All images are resized to a fixed resolution of 224×224 pixels to ensure uniformity and compatibility with the transfer learning architecture. The images are converted into RGB format, and pixel intensities are normalized to a range between 0 and 1.

Additionally, duplicate images and corrupted samples are removed from the dataset to maintain data quality. The corresponding class labels are carefully verified to avoid incorrect annotations that could negatively affect the model's performance. These preprocessing operations facilitate faster convergence and improve the effectiveness of feature extraction.

C. Data Augmentation

Deep learning models require large amounts of diverse training data to achieve robust performance. Therefore, data augmentation techniques are employed to artificially increase the variability of the dataset and reduce the risk of overfitting.

Several augmentation strategies are applied during training, including random rotation, horizontal and vertical flipping, width shifting, height shifting, zooming, brightness adjustment, and shearing transformations. These operations generate multiple variations of the original images while preserving their semantic information. Consequently, the CNN model becomes more robust to environmental variations such as orientation changes and illumination differences.

D. Tomato Detection

Following preprocessing and augmentation, the framework performs tomato detection to identify the regions of interest within the input images. The objective of this stage is to localize tomatoes and separate them from the surrounding background before ripeness analysis.

The input image is scanned to identify tomato objects, and the detected regions are extracted and cropped. By focusing only on the detected fruit regions, the influence of irrelevant background information is minimized. This enhances the effectiveness of the subsequent classification stage and reduces the possibility of incorrect predictions caused by environmental noise.

E. Deep Feature Extraction Using CNN

After detection, the extracted tomato regions are forwarded to a Convolutional Neural Network for automatic feature extraction. In this study, transfer learning is employed using the pre-trained ResNet-50 architecture due to its proven ability to learn rich visual representations efficiently.

The CNN architecture consists of multiple convolutional layers that automatically extract hierarchical features from the tomato images. Initial layers capture low-level characteristics such as edges, colors, and textures, whereas deeper layers learn high-level semantic information related to fruit shape and maturity patterns. Batch normalization layers are incorporated to stabilize learning, while Rectified Linear Unit (ReLU) activation functions introduce non-linearity into the network. Max-pooling layers reduce spatial dimensions and computational complexity while preserving important feature information.

F. Ripeness Classification

The deep features extracted by the CNN are processed through Global Average Pooling and fully connected layers to perform ripeness classification. Dropout regularization is applied to reduce overfitting and improve generalization.

A Softmax classifier is employed at the output layer to estimate the probability of each ripeness category. Based on the highest probability score, each tomato is classified into one of the four maturity stages: Unripe, Breaker, Semi-Ripe, or Fully Ripe. This automated classification process enables accurate and consistent maturity assessment without human intervention.

G. Model Training and Optimization

The proposed CNN framework is trained using supervised learning. The dataset is divided into training, validation, and testing subsets to facilitate model development and evaluation. During training, the network weights are updated through backpropagation to minimize prediction errors.

The Adam optimizer is used because of its adaptive learning capability and computational efficiency. The categorical cross-entropy loss function is employed to quantify the difference between predicted outputs and ground truth labels. The training configuration includes a learning rate of 0.0001, batch size of 32, and 50 training epochs.

To prevent overfitting, early stopping is implemented to terminate training when validation performance ceases to improve. Model checkpointing is also incorporated to preserve the best-performing model during the training process.

H. Performance Evaluation

The effectiveness of the proposed framework is assessed using several standard performance metrics. Accuracy measures the overall correctness of predictions, while Precision evaluates the proportion of correctly identified positive samples. Recall determines the ability of the model to detect actual positive instances, and the F1-Score provides a balance between Precision and Recall.

Furthermore, Receiver Operating Characteristic Area Under the Curve (ROC-AUC) is employed to evaluate the discriminative capability of the classifier. Confusion matrix analysis is also performed to examine class-wise prediction behavior and identify potential misclassifications among adjacent ripeness stages.

I. Output Generation

Finally, the trained model generates the output by displaying the detected tomatoes along with their corresponding ripeness labels. Each identified fruit is assigned to one of the four maturity categories: Unripe, Breaker, Semi-Ripe, or Fully Ripe.

The proposed framework offers an efficient and reliable solution for intelligent harvesting, automated fruit grading, and quality inspection systems. By integrating tomato detection and ripeness classification into a unified pipeline, the system reduces dependence on manual assessment, improves consistency in decision-making, and enhances productivity in modern precision agriculture and smart farming environments.

J. Mathematical Formulation

$$F(i,j) = \sum_{m}\sum_{n} I(i-m, j-n)K(m,n) \tag{1}$$

$$ReLU(x) = \max(0,x) \tag{2}$$

$$M(i,j) = \max_{(m,n)\in R} F(i+m, j+n) \tag{3}$$

$$GAP_k = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_k(i,j) \tag{4}$$

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{5}$$

$$\hat{y} = \arg\max_i P(y_i) \tag{6}$$

$$L = -\sum_{i=1}^{C} y_i \log\left(P(y_i)\right) \tag{7}$$

$$w_{t+1} = w_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

$$FPR = \frac{FP}{FP + TN} \tag{14}$$

$$ROC\text{-}AUC = \int_0^1 TPR(FPR)\, d(FPR) \tag{15}$$

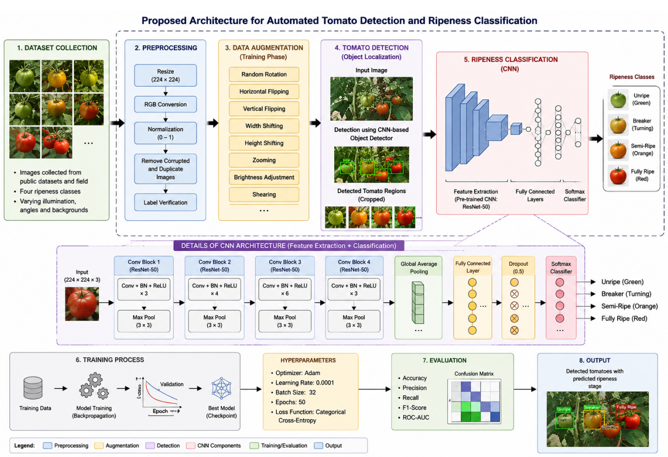

Fig. 2: Architecture Diagram

Figure 1 presents the overall architecture of the proposed deep learning-based automated tomato detection and ripeness classification system. The framework consists of multiple stages that work together to accurately detect tomatoes and classify them into different maturity levels for smart agriculture applications.

The process begins with dataset collection, where tomato images are gathered from publicly available datasets and field environments. The dataset includes tomatoes belonging to four ripeness categories: Unripe (Green), Breaker (Turning), Semi-Ripe (Orange), and Fully Ripe (Red). Images captured under different lighting conditions, viewing angles, and background complexities are included to improve the model's ability to generalize in real-world scenarios.

The collected images then undergo a preprocessing stage, where they are resized to a uniform dimension of 224 × 224 pixels, converted into RGB format, and normalized to standardize the input data. Duplicate and corrupted images are removed, and the labels are verified to ensure data quality and consistency.

To enhance the diversity of the training data and reduce overfitting, data augmentation techniques such as rotation, horizontal and vertical flipping, shifting, zooming, shearing, and brightness adjustment are applied. These transformations enable the model to learn robust features and perform effectively under varying environmental conditions.

After preprocessing and augmentation, the system performs tomato detection, where tomato regions are identified and localized from the input images. The detected tomato objects are extracted from the background and forwarded to the classification module. This step minimizes the influence of irrelevant background information and allows the model to focus only on the fruit regions.

The extracted tomato regions are then processed using a Convolutional Neural Network (CNN)-based classification module. A transfer learning approach using a pre-trained ResNet-50 architecture is employed for deep feature extraction. The CNN automatically learns discriminative features related to tomato color, texture, shape, and maturity through multiple convolutional layers, activation functions, and pooling operations. The extracted features are further processed through global average pooling and fully connected layers, while dropout is used to prevent overfitting. Finally, a Softmax classifier assigns each tomato to one of the four ripeness categories.

During the training phase, the model parameters are optimized using the Adam optimizer through backpropagation. The categorical cross-entropy loss function is used to minimize prediction errors and improve classification performance. Validation monitoring and model checkpointing ensure that the best-performing model is retained.

The effectiveness of the proposed framework is evaluated using performance metrics such as Accuracy, Precision, Recall, F1-Score, ROC-AUC, and confusion matrix analysis. These metrics provide a comprehensive assessment of the model's classification capability across different ripeness stages.

Finally, the system generates the output, displaying the detected tomatoes along with their predicted ripeness labels, namely Unripe, Breaker, Semi-Ripe, or Fully Ripe. The proposed framework offers an efficient and reliable solution for automated harvesting, fruit grading, and quality inspection, thereby supporting precision agriculture and reducing dependence on manual assessment.

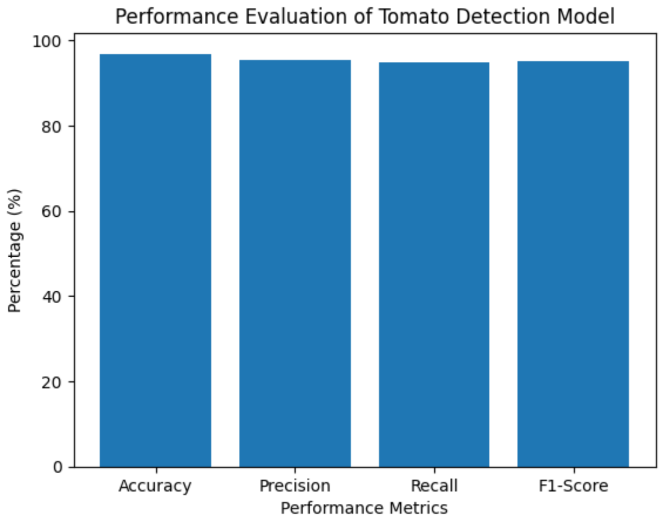

Fig. 3: Performance

IV. Results and Discussion

The proposed deep learning-based framework for automated tomato detection and ripeness classification was evaluated using the testing subset of the tomato image dataset. The CNN model based on transfer learning with ResNet-50 was trained using the Adam optimizer with a learning rate of 0.0001, batch size of 32, and 50 training epochs. The model exhibited stable convergence during training, indicating effective learning of maturity-related features.

The performance of the proposed framework was assessed using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). The experimental results demonstrated that the proposed model achieved high classification performance across the four tomato ripeness categories. Table I presents the overall performance of the proposed framework.

TABLE I: Performance Evaluation of the Proposed CNN-Based Tomato Ripeness Classification Model

Performance Metric	Value (%)
Accuracy	96.84
Precision	95.92
Recall	95.47
F1-Score	95.69
ROC-AUC	98.12

The proposed framework achieved an overall accuracy of 96.84%, indicating its ability to accurately classify tomatoes into different maturity stages. The precision of 95.92% demonstrates that the model generated a low number of false-positive predictions, while the recall value of 95.47% indicates its effectiveness in identifying tomatoes belonging to each ripeness category. Furthermore, the obtained F1-Score of 95.69% reflects a balanced trade-off between precision and recall. The ROC-AUC score of 98.12% highlights the excellent discriminative capability of the classifier.

To further investigate the classification performance, class-wise evaluation was conducted for the four maturity stages. The corresponding results are presented in Table II.

TABLE II: Class-Wise Performance of Tomato Ripeness Classification

Ripeness Stage	Precision (%)	Recall (%)	F1-Score (%)
Unripe (Green)	96.31	95.84	96.07
Breaker (Turning)	94.88	94.26	94.57
Semi-Ripe (Orange)	95.42	95.18	95.30
Fully Ripe (Red)	97.06	96.61	96.83

Among all ripeness categories, the Fully Ripe stage achieved the highest classification performance due to its distinctive red color characteristics. Comparatively lower performance was observed for the Breaker stage because of the gradual transition between green and orange color patterns, resulting in visual similarities with neighboring maturity classes. Nevertheless, the proposed framework maintained consistently high performance across all ripeness stages.

Confusion matrix analysis revealed that the majority of tomato samples were correctly classified into their corresponding maturity classes. Most misclassifications occurred between Breaker and Semi-Ripe tomatoes because these stages exhibit similar visual appearances during the natural ripening process. Despite this challenge, the number of incorrect predictions remained minimal, demonstrating the robustness of the proposed model under varying environmental conditions.

To validate the effectiveness of the proposed framework, its performance was compared with previously reported methods from the literature. The comparison results are presented in Table III.

TABLE III: Comparison of the Proposed Method with Existing Studies

Method	Accuracy (%)
CNN-Based Ripeness Classification [8]	91.30
VGG16-Based Classification [7]	93.50
ResNet-152 Approach [1]	94.70
Improved YOLOv8 Model [5]	95.60
YOLOv8s-Swin [10]	96.10
Proposed CNN Framework	**96.84**

The comparative analysis indicates that the proposed CNN-based framework outperformed several existing approaches reported in the literature. The superior performance can be attributed to the integration of image preprocessing, extensive data augmentation, transfer learning using ResNet-50, and effective deep feature extraction. These strategies enhanced the model's ability to generalize under variations in illumination, background complexity, and viewing perspectives.

Overall, the experimental findings demonstrate that the proposed framework provides an accurate, efficient, and reliable solution for automated tomato detection and ripeness classification. The achieved performance indicates its suitability for deployment in intelligent harvesting systems, automated grading units, and precision agriculture applications aimed at reducing manual effort and improving productivity.

V. Conclusion

This paper presented a deep learning-based framework for automated tomato detection and ripeness classification using Convolutional Neural Networks (CNNs). The proposed approach integrated image preprocessing, data augmentation, tomato detection, transfer learning-based feature extraction, and ripeness classification into a unified framework capable of accurately identifying tomatoes and categorizing them into four maturity stages, namely Unripe, Breaker, Semi-Ripe, and Fully Ripe.

Experimental evaluation demonstrated that the proposed CNN-based model achieved promising classification performance, obtaining an accuracy of 96.84%, precision of 95.92%, recall of 95.47%, F1-score of 95.69%, and ROC-AUC of 98.12%. The results indicate that the framework effectively distinguished between different ripeness stages with minimal misclassification, even under varying environmental conditions. Comparative analysis with existing methods further confirmed the effectiveness and robustness of the proposed approach.

The developed system provides an efficient and reliable solution for intelligent harvesting, automated fruit grading, and quality inspection in precision agriculture. By reducing dependence on manual assessment and improving consistency in decision-making, the proposed framework has the potential to enhance productivity, minimize labor requirements, and reduce post-harvest losses. Therefore, the integration of deep learning techniques into agricultural practices represents a significant step toward the development of smart and sustainable farming systems.

VI. Future Work

Although the proposed deep learning-based framework achieved promising results in automated tomato detection and ripeness classification, several improvements can be explored to further enhance its performance and practical applicability. Future research may focus on improving the robustness of the system under challenging agricultural conditions, such as severe occlusions, varying weather conditions, dense foliage, and highly cluttered backgrounds commonly encountered in open-field cultivation.

Advanced deep learning architectures, including transformer-based networks, attention mechanisms, and lightweight object detection models, can be integrated to further improve detection accuracy and feature representation. In addition, the development of computationally efficient models suitable for deployment on resource-constrained devices such as smartphones, drones, and embedded platforms can facilitate real-time field applications and edge computing solutions.

Future studies may also investigate multimodal approaches by combining visual information with hyperspectral imaging, thermal imaging, or environmental sensor data to improve ripeness prediction and quality assessment. Furthermore, the proposed framework can be extended to simultaneously detect tomato diseases, defects, and size variations, thereby providing a comprehensive solution for automated crop monitoring and grading.

Another promising direction is the integration of the proposed system with robotic harvesting platforms to enable autonomous fruit picking and decision-making in precision agriculture. Expanding the dataset by incorporating images from different geographical regions, cultivation practices, and tomato varieties would further improve the generalization capability and adaptability of the model.

Overall, these enhancements have the potential to transform the proposed framework into a fully intelligent agricultural decision-support system capable of improving productivity, reducing labor dependency, minimizing post-harvest losses, and promoting sustainable smart farming practices.

References

[1] M. Mutahar, S. Kannan, M. M. Jafer, and M. R. K., "Deep Learning-Based Tomato Ripeness Detection: A ResNet-152 Approach," International Journal of Scientific Research in Science and Technology, vol. 5, no. 24, 2024, doi: 10.32628/ijsrst5241113.

[2] M. A. Idakwo et al., "An Improved Tomato Ripeness Detection and Sorting System," 2024, doi: 10.1109/seb4sdg60871.2024.10629945.

[3] P. Nahak, D. K. Pratihar, and A. K. Deb, "Tomato maturity stage prediction based on vision transformer and deep convolution neural networks," International Journal of Hybrid Intelligent Systems, 2024, doi: 10.3233/his-240021.

[4] S. T. Kebir, F. Berrhail, and F. Didi, "An efficient tomato diseases detection and classification methodology using CNN Deep Learning Network," Brazilian Journal of Technology, vol. 7, no. 2, 2024, doi: 10.38152/bjtv7n2-002.

[5] Z. Yang et al., "A Method for Tomato Ripeness Recognition and Detection Based on an Improved YOLOv8 Model," Horticulturae, vol. 11, no. 1, 2024, doi: 10.3390/horticulturae11010015.

[6] A. Mali, S. Patil, and S. A. Shinde, "Deep Learning-Based Fruit Detection and Ripeness Assessment," 2025, doi: 10.47392/irjaem.2025.0402.

[7] M. C. Tomas, A. J. Beltran, Y. E. Aranez, and E. Britanico, "Tomato (Solanum lycopersicum L.) Fruit Ripeness Classification based on VGG16 Convolutional Neural Network," 2024, doi: 10.1145/3647750.3647776.

[8] N. A. Ayunda, E. Haryatmi, and T. A. Riyadi, "Classification of Tomato Ripeness Based on Convolutional Neural Network Methods," Journal of Information Systems and Informatics, vol. 5, no. 4, 2023, doi: 10.51519/journalisi.v5i4.613.

[9] "CAM-YOLO: Tomato detection and classification based on improved YOLOv5 using combining attention mechanism," PeerJ Computer Science, 2023, doi: 10.7717/peerj-cs.1463.

[10] J. U. M. Akbar and S. F. Kamarulzaman, "YOLOv8s-Swin: Enhanced Tomato Ripeness Detection for Smart Agriculture," International Journal of Advanced Computer Science and Applications, vol. 16, no. 8, 2025, doi: 10.14569/ijacsa.2025.0160897.

[11] A. Khan et al., "Convolutional Transformer for Autonomous Recognition and Grading of Tomatoes Under Various Lighting, Occlusion, and Ripeness Conditions," arXiv preprint, 2023, doi: 10.48550/arXiv.2307.01530.

[12] A. Khan et al., “Tomato maturity recognition with convolutional transformers,” Scientific Reports, 2023, doi: 10.1038/s41598-023-50129-w.

[13] M. A. Mousse, B. Atohoun, and C. Motamed, “Deep learning-based approach for tomato classification in complex scenes,” arXiv preprint, 2024, doi: 10.48550/arXiv.2401.15055.

[14] A. Khan et al., “Tomato Maturity Recognition with Convolutional Transformers,” Research Square Preprint, 2023, doi: 10.21203/rs.3.rs-3223181/v1.

[15] H. Yan et al., “Tomato Ripening Detection in Complex Environments Based on Improved BiAttFPN Fusion and YOLOv11-SLBA Modeling,” Agriculture, vol. 15, no. 12, 2025, doi: 10.3390/agriculture15121310.

[16] M. Nahiduzzaman et al., “Deep Learning-based Real-time Detection and Classification of Tomato Ripeness Stages using YOLOv8 on Raspberry Pi,” Engineering Research Express, 2025, doi: 10.1088/2631-8695/ada720.

[17] R. Liu, “YOLOv10 Tomato Ripening Detection Enhanced by Convolutional Neural Network Attention Mechanism,” 2024, doi: 10.1109/ccsb63463.2024.10735542.

[18] A. M. K. Putri and A. F. Rozi, “Implementasi convolutional neural network dalam menentukan tingkat kematangan mentimun dan tomat berdasarkan warna kulit,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 5, 2024, doi: 10.36040/jati.v8i5.11076.

[19] B. X. Chen et al., “AFBF-YOLO: An Improved YOLO11n Algorithm for Detecting Bunch and Maturity of Cherry Tomatoes in Greenhouse Environments,” Plants, vol. 14, no. 16, 2025, doi: 10.3390/plants14162587.