



Computer Vision and Field Validation of an Artificial Intelligence-Based Tomato Grader for Large-Scale Production in Northeastern Thailand

Khunnithi Doungpueng¹, Kittikun Pituprompan¹, Chainarong Lomchangkum², Kantapon Premprayoon³, Thanaporn Singhpoo⁴ and Anuwat Pachanawan^{5,*}

¹ Department of Mechanical Technical Education, Faculty of Technical Education, Rajamangala University of Technology Isan Khon Kaen Campus, Khon Kaen 40000, Thailand.

² Department of Food and Biological Engineering, Faculty of Engineering, Rajamangala University of Technology Isan Khon Kaen Campus, Khon Kaen 40000, Thailand

³ Department of Agricultural Machinery Engineering, Faculty of Engineering, Rajamangala University of Technology Isan, Khon Kaen Campus, Khon Kaen, 40000, Thailand

⁴ Department of Mechanical Engineering, Faculty of Engineering, Rajamangala University of Technology Isan Khon Kaen Campus, Khon Kaen 40000, Thailand

⁵ Department of Agricultural Machinery Engineering, Faculty of Engineering and Industrial Technology, Kalasin University, Kalasin 46000, Thailand

*Corresponding author: anuwat.pa@ksu.ac.th

Received 4 November 2025

Revised 12 March 2026

Accepted 20 March 2026

Abstract

Tomato (*Solanum lycopersicum L.*) quality grading based on visual inspection often yields inconsistent results and reduces market value. The 'Perfect Gold 111' variety presents distinct morphological traits, including a characteristic green-to-red color transition, specific calyx structure, and defect patterns such as greenback distribution and suberization. These characteristics differ substantially from internationally studied cultivars, rendering generic pre-trained models insufficient for accurate grading under the Thai TACFS 1503–2007 standard. This study developed an automated grading model for 'Perfect Gold 111' tomatoes using a deep learning model based on a flow-based (node-based) architecture integrated with the Robot Operating System (ROS) framework and implemented on the CiRA CORE platform. A total of 220 samples were collected and graded according to the TACFS 1503–2007 standard. Top and side-view images were used to create a dataset comprising 165 tomatoes for training and 55 for testing. Model performance was evaluated using Precision, Recall, F1-Score, and Accuracy, and was compared with manual grading performed by farmers. The AI model achieved 80.00% of accuracy, outperforming farmer grading, which achieved 52.72% accuracy. In addition, the model reduced misclassification among visually similar grades and provided consistent, quantitative assessments of color, shape, and defects. These findings highlight the potential of AI-based grading systems to improve quality consistency, reduce labor, and support automated postharvest sorting for both smallholder and industrial tomato production.

Keywords: Tomato grading, Deep learning, Computer vision, Quality classification, Postharvest automation

1. Introduction

Tomato (*Solanum lycopersicum L.*) is one of the most widely cultivated vegetable crops worldwide and is valued for both fresh consumption and industrial processing [1]. In Thailand, tomato production plays a significant role in the agricultural economy, particularly in the northeastern region, where large-scale cultivation is concentrated [2-3]. The 'Perfect Gold 111' variety is widely cultivated by farmers in Nong Khai Province. Typically, tomato grading relies on manual visual inspection based on farmers' experience, which often results in inconsistent quality classification and reduced market value [3].

Artificial intelligence (AI) has recently emerged as a key technology in modern agriculture, supporting automation and precision in agricultural production systems [5]. Machine learning and computer vision enable the recognition, detection, and classification of agricultural products with high accuracy and repeatability [6-7]. Several studies have successfully applied AI-based image processing for agricultural classification tasks, such as rice grain sorting and rubber sheet inspection, demonstrating performance comparable to that of human evaluation [8-11].

Within the field of tomato quality assessment, significant research has recently focused on AI-based grading and sorting systems. Among the most widely adopted architectures, DenseNet-201 with transfer learning [16] and ensemble classifier pipelines [20] have shown reliable discrimination based on color and surface texture. In addition, shape-based classification [17] and CNN-driven defect detection [18] have further extended the scope of automated assessment, while fully integrated machine vision systems that jointly analyze color, size, and shape have been successfully implemented for postharvest sorting [19]. YOLOv9 has also been applied to real-time ripeness detection [21], and a recent comprehensive review [22] confirmed the substantial progress achieved through sensor-based and machine learning approaches. Despite these advances, existing studies exhibit several limitations in the context of Thai agricultural production: (1) most models are predominantly developed and validated using international tomato cultivars whose morphological and defect characteristics differ from local varieties; (2) current systems do not align with the Thai national grading standard (TACFS 1503–2007), which defines class boundaries based on specific defect thresholds that differ from international criteria; and (3) field validation under actual smallholder farming conditions in Thailand has not yet been reported. These gaps underscore the need for a cultivar-specific, standard-compliant AI grading model developed under local agricultural conditions.

Consequently, AI application for the ‘Perfect Gold 111’ variety remains limited. This variety exhibits unique morphological characteristics, including a distinctive green-to-deep-red color gradient during ripening, a prominent calyx structure, and variety-specific surface defect patterns — such as greenback discoloration around the stem and suberization at the calyx end — that do not conform to the feature distributions captured by off-the-shelf models. Furthermore, quality grading in this study strictly follows the TACFS 1503–2007 national standard, which defines class boundaries based on defect thresholds (e.g., suberization area $\leq 1 \text{ cm}^2$ and crack length $\leq 1 \text{ cm}$ for Class I) that differ substantially from international grading criteria used to train generic models. Collectively, these variety-specific and standard-specific factors necessitate the development of a custom deep learning model trained on locally sourced samples under Thai agricultural conditions.

Therefore, this study developed a deep learning-based classification model for grading ‘Perfect Gold 111’ tomatoes in accordance with the TACFS 1503–2007 national standard, with a target classification accuracy of at least 75% across all quality classes. Four standard performance metrics — Precision, Recall, F1-Score, and Accuracy — were employed to compare model output against manual quality assessments conducted by experienced farmers. The CiRA CORE platform, a collaborative initiative among King Mongkut’s Institute of Technology Ladkrabang, Khon Kaen University, and King Mongkut’s University of Technology North Bangkok, was adopted for model development because of its suitability for practical agricultural deployment. The platform operates on standard computing hardware, thereby substantially reducing the technical and financial barriers to adoption for smallholder farming communities in rural Thailand.

2. Materials and methods

This study was conducted in three sequential phases: (1) tomato quality grading according to the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007) [12], (2) development of a grading model for the ‘Perfect Gold 111’ tomato variety, and (3) testing and evaluation of the developed grading model. The detailed procedures are as follows:

2.1 Quality grading according to the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007)

Samples of the ‘Perfect Gold 111’ tomato variety, cultivated for three months, were collected from a large-scale tomato farming group in Phan Phrao Subdistrict, Si Chiang Mai District, Nong Khai Province. Harvesting was performed by the farmers, and both green and red tomatoes were placed into collection containers with a capacity of 25 kg (Figure 1). The researchers then randomly sampled only the red tomatoes for analysis. Each sampled tomato was individually labeled with a unique identification number and graded according to the quality-classification requirements specified in the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007) [12]. The classification criteria are summarized in Figure 2.

(1) Extra Class

Tomatoes in this class represent the highest quality. They must conform to varietal characteristics in shape, size, and development, and possess appropriate flesh firmness. Tomato color must be uniform and maturity level

must fall within the acceptable range defined by the standard. A small amount of green coloration around the stem (greenback) is acceptable if it is characteristic of the variety and does not impair overall quality. Tomatoes must be intact, free from bruising or damage, and capable of withstanding transportation and storage. These tomatoes are suitable for experimental use or packaging.

(2) Class I

Tomatoes in this class are of good quality and possess firm flesh consistent with varietal traits. Tomatoes classified as Class I should be of uniform size and should not exhibit abnormal green areas (greenback). Minimal green coloration near the stem is permitted if it is characteristic of the variety and does not affect the product's appearance. Acceptable defects and limits include:

1. Slight healed scars or cracks with a length not exceeding 1 cm.
2. Absence of excessive protuberances.
3. A small umbilicus (small calyx-end scar) without suberization.
4. Areas of suberization not exceeding 1 cm².
5. Linear scars are not exceeding two-thirds of the tomato diameter.

(3) Class II

This class includes tomatoes that do not meet the requirements for Extra Class or Class I. Characteristics include:

1. More pronounced protuberances than those permitted in Class I, provided that they do not cause deformity.
2. Presence of a natural calyx-end scar (one umbilicus).
3. Suberization at the calyx end not exceeding 2 cm².
4. Thin longitudinal blossom scars (seam-like).

Tomatoes with quality lower than that specified for Class II are classified as Class III (rejected/lowest grade).

2.2 Measurement and data recording of physical properties

For each sampled tomato, the following physical properties were measured and recorded to support quality classification and subsequent model development:

1. Area of greenback (green coloration around the stem).
2. Length of shallow scars or cracks.
3. Number of protuberances projecting on the tomato surface.
4. Suberization area (healed corky tissue).
5. Length of linear scars.
6. Tomato diameter.
7. Tomato weight.

After measurements were completed, the researchers explained the standard grading criteria to the farmers and presented representative tomato samples for each quality class. The farmers then independently graded the sampled tomatoes, and their grading results were recorded for comparison with the researcher-assigned classifications.

(A)



(B)



Figure 1 Sampling process for the ‘Perfect Gold 111’ tomato variety: (A) harvesting performed by farmers and (B) random sampling and labeling of the selected tomatoes.

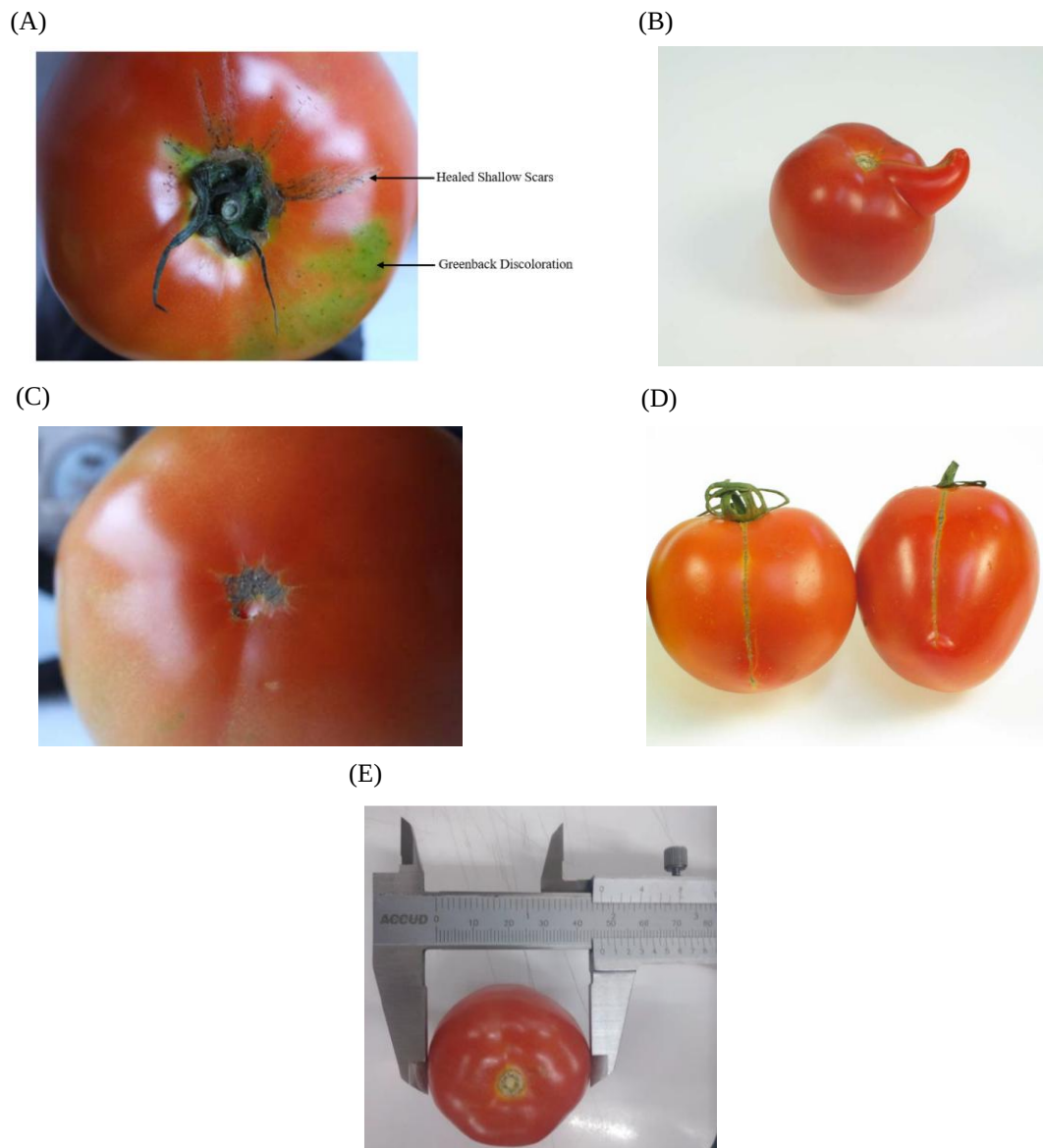


Figure 2 Examples of physical characteristic measurement of the ‘Perfect Gold 111’ tomato used for quality classification according to the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007) [12]: (A) greenback area and shallow scar length, (B) protuberances on the tomato surface, (C) suberization area, (D) length of linear scar, and (E) tomato diameter.

2.3 Model Development for Quality Classification of ‘Perfect Gold 111’ Tomatoes

The model development process for quality classification of the ‘Perfect Gold 111’ tomato variety began with data collection and preparation. Images of tomatoes randomly sampled from the previous phase were captured from both top and side views. The images were captured to clearly reveal visible surface defects on the tomatoes, which served as essential information for model training and development (Figure 3). It should be noted that the deep learning model was developed exclusively using image data. The seven physical measurements recorded in Section 2.2 (greenback area, crack length, number of protuberances, suberization area, linear scar length, tomato diameter, and tomato weight) were used solely for researcher-based and farmer-based quality grading in accordance with the TACFS 1503–2007 standard and were not included as input features in the model. The model therefore operated entirely on pixel-level visual information extracted from the top and side view images.

To ensure reproducibility of the image acquisition process, all photographs were captured under standardized environmental conditions using a fixed imaging setup. Images were captured using a Canon EOS 700D camera at a resolution of 2592 x 1728 pixels and saved in JPEG (.jpg) format. Illumination was provided by LED lighting arranged to ensure uniform and consistent light distribution across the tomato surface, with no direct shadows or specular reflections observed during image acquisition. Each tomato was placed individually at the center of a matte black cardboard background to minimize surface reflectance and provide high visual contrast for defect detection. Top-view images were acquired with the camera positioned directly overhead at 90° and a fixed lens-

to-surface distance of 32 cm, whereas side-view images were captured horizontally at 0° at the same distance from the tomato's equatorial plane. Camera position and lighting conditions were maintained constant throughout the imaging process to limit variability in illumination and perspective. No post-processing filters or color correction were applied, thereby preserving the raw visual characteristics of the samples as encountered under field conditions.

The acquired images were organized into a dataset for model development. This dataset was divided into two subsets: (1) training data and (2) testing data. The training and testing sets were comprised of 75% and 25% of the total dataset, respectively.

In this study, the model was developed using the CiRA CORE platform, an artificial intelligence platform jointly developed by King Mongkut's Institute of Technology Ladkrabang, Khon Kaen University, and King Mongkut's University of Technology North Bangkok. The CiRA CORE platform enables the development of deep learning models for visual recognition of tomato physical characteristics from training images, as illustrated in Figure 4.

To address the relatively limited dataset size and reduce the risk of overfitting, several data augmentation techniques were applied to the training images during model development within the CiRA CORE platform. These techniques included random horizontal flipping, random vertical flipping, rotation across a range of angles, brightness adjustment, and contrast variation. Collectively, these augmentation methods were applied to simulate variations in tomato orientation, illumination, and surface appearance during image acquisition, thereby expanding dataset diversity without the need for additional physical samples. The CiRA CORE platform further incorporates internal regularization mechanisms within its deep learning framework. The combined effects of augmentation and regularization improved the model's ability to generalize across variations in lighting, surface texture, and viewing angle.

A dataset comprising 220 tomatoes was considered appropriate for the exploratory scope of this work, which focused on a specific Thai cultivar under a defined national standard. Previous studies have shown that models trained on datasets of comparable size can achieve acceptable performance when augmentation is applied and the classification task is clearly defined [27, 29], thereby supporting the suitability of the present sample size for a pilot-scale investigation.

Upon completion of the training process, a deep learning model for tomato quality classification was obtained and subsequently applied for testing and performance evaluation in the next phase.

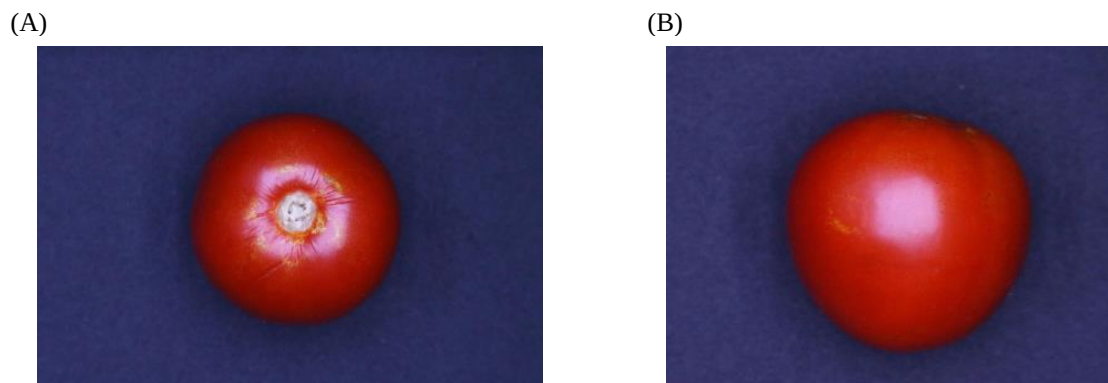


Figure 3 Representative acquired image of the 'Perfect Gold 111' tomato: (A) top view image and (B) side view image.

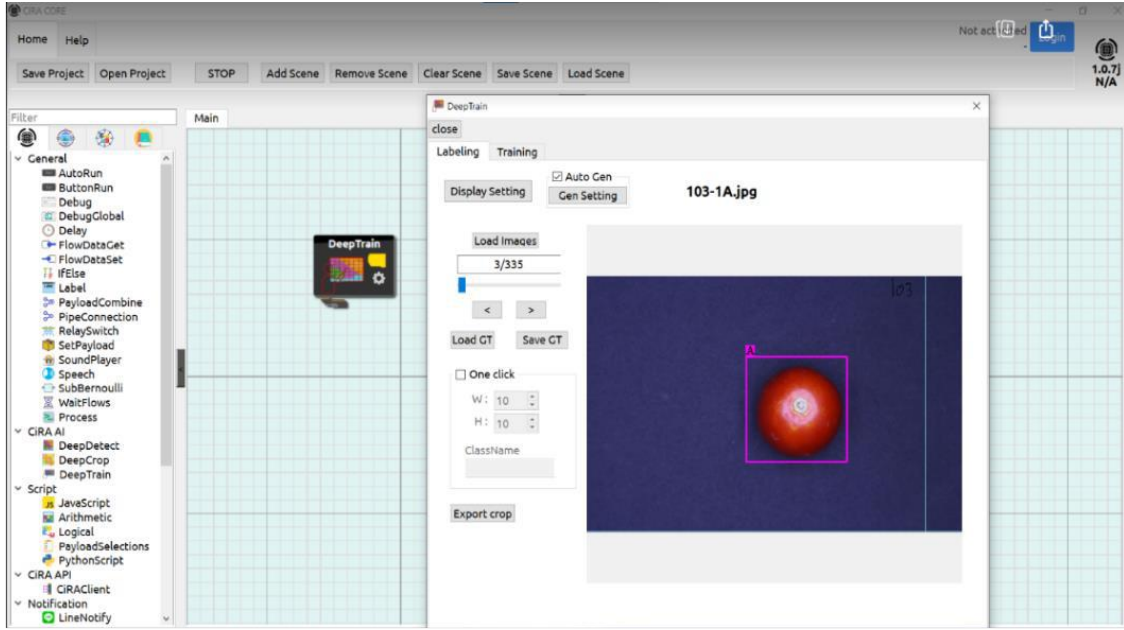


Figure 4 Labeling process for tomato samples for model development.

2.4 Testing and Evaluation of the Developed Tomato Quality Classification Model

The performance of the developed model for classifying the quality grades of the ‘Perfect Gold 111’ tomato was evaluated using the reserved 25% testing subset. The model efficacy was assessed using four standard metrics: Precision, Recall, F1-Score [13-15], and Accuracy. The dataset was partitioned using a single holdout split of 75% training data and 25% testing data, which is a standard evaluation approach in applied agricultural AI studies [27, 29]. It is acknowledged that k-fold cross-validation would provide a more statistically robust estimate of model generalizability, particularly given the limited dataset size. However, due to constraints within the current training pipeline of the CiRA CORE platform, k-fold cross-validation was not implemented in this study. This limitation is acknowledged in the present work, and the application of k-fold cross-validation is recommended for future studies using expanded datasets.

(1) Precision

Precision represents the ratio of correctly predicted positive samples to the total number of positive predictions, as expressed in Equation (1):

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

where TP (True Positive) denotes positive samples correctly classified as belonging to a given class (actual positive, predicted positive), whereas FP (False Positive) denotes negative samples incorrectly classified as belonging to that class (actual negative, predicted positive).

(2) Recall

Recall measures the proportion of actual positive samples that are correctly classified, as given in Equation (2):

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

where FN (False Negative) denotes positive samples incorrectly classified as belonging to the true class (actual positive, predicted negative).

(3) F1-Score

The F1-Score represents the harmonic mean of Precision and Recall, providing a single metric for evaluating model performance, as shown in Equation (3):

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

(4) Accuracy

Accuracy indicates the overall correctness of the model predictions and is expressed in Equation (4):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

where TN (True Negative) represents negative samples correctly identified as not belonging to a given class (actual negative, predicted negative).

3. Results and discussion

3.1 Results of Tomato Quality Classification Based on the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007)

Tomato quality classification based on the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007) [12] was conducted using a total of 220 tomato samples. Classification performed by the researchers revealed the following distribution: 56 tomatoes were classified as Extra Class, 123 tomatoes as Class I, 28 tomatoes as Class II, and 13 tomatoes as Class III. Meanwhile, the classification performed by the farmers yielded slightly different results: 83 tomatoes were categorized as Extra Class, 71 tomatoes as Class I, 60 tomatoes as Class II, and 6 tomatoes as Class III. The comparative results of tomato quality classification conducted by researchers and farmers are summarized in Table 1.

Table 1 Comparative results of tomato quality classification performed by researchers and farmers according to the National Bureau of Agricultural Commodity and Food Standards (TACFS 1503–2007).

Quality class	The number of tomatoes classified by	
	Researcher (tomatoes)	Farmer (tomatoes)
Extra Class	56	83
Class I	123	71
Class II	28	60
Class III	13	6
Total	220	220

3.2 Results of Testing and Evaluation of the Tomato Quality Classification Model

A total of 220 tomatoes classified by the researchers, as described in Section 2.1, were used for testing and evaluation of the developed model. The dataset was divided into two subsets: training data, which consisted of 165 tomatoes (representing 75% of the total samples) and the testing data (55 tomatoes, representing 25% of the total samples).

After dataset separation, images of the tomatoes were captured from two perspectives: top view and side view. Upon completion of image acquisition, the collected images were used for model training and testing—330 images for training and 110 images for testing. The dataset composition is summarized in Table 2.

Table 2 Distribution of tomato samples across quality classes for the training and testing datasets.

Quality class	Number of tomatoes (Training Data)	Number of images (Training Data)	Number of tomatoes (Testing Data)	Number of images (Testing Data)
Extra Class	42	84	14	28
Class I	93	186	30	60
Class II	20	40	8	16
Class III	10	20	3	6
Total	165	330	55	110

Table 3 shows the confusion matrix for tomato quality classification performed by farmers. The results revealed that the farmers correctly classified 9 tomatoes in the Extra Class, 15 tomatoes in Class I, 4 tomatoes in Class II, and 1 tomato in Class III. Evaluation of the farmers' classification performance using Precision, Recall, and F1-score revealed that the corresponding values for Extra Class were 43.00%, 64.00%, and 51.40%, respectively. For the Class I, the Precision, Recall, and F1-Score were 65.00%, 48.00%, and 55.22%, respectively. For Class II, the corresponding values were 40.00%, 57.00%, and 47.42%, respectively. For Class III, the Precision, Recall, and F1-Score were 100.00%, 33.00%, and 49.62%, respectively. Overall, the farmers' tomato quality classification achieved an accuracy of 52.72% across all quality classes, as shown in Table 4.

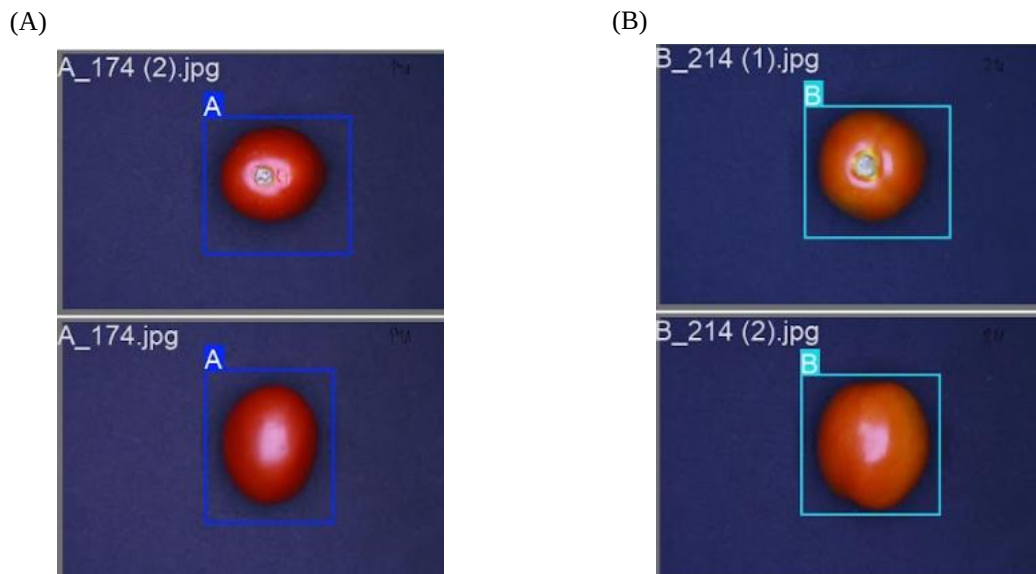
Table 3 Confusion matrix of tomato quality classification for ‘Perfect Gold 111’ tomato variety performed by farmers.

Predicted class (by farmers)	Extra Class	Class I	Class II	Class III	Total
Extra Class	9	12	0	0	21
Class I	5	15	3	0	23
Class II	0	4	4	2	10
Class III	0	0	0	1	1
Total	14	31	7	3	55

Table 4 Performance evaluation of tomato quality classification conducted by farmers

Quality class	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Extra Class	43.00	64.00	51.40	
Class I	65.00	48.00	55.22	
Class II	40.00	57.00	47.42	
Class III	100.00	33.00	49.62	
Overall				52.72

Figure 5 illustrates examples of tomato quality classification generated by the developed model, while Table 5 presents the confusion matrix of the classification results. The findings revealed that the developed model correctly classified 11 tomatoes in the Extra Class, 26 tomatoes in Class I, 5 tomatoes in Class II, and 2 tomatoes in Class III. Class-level evaluation of the developed model indicated strong performance across all quality grades. The Extra Class and Class I categories achieved F1-Scores of 75.66% and 83.92%, respectively, with closely balanced Precision and Recall values (73%/79% and 84%/84%, respectively), reflecting reliable discrimination at these class boundaries. Class II attained a Precision of 83% and a Recall of 71% (F1-Score: 77.00%), while Class III recorded 66% for both metrics (F1-Score: 65.91%), the latter partially attributable to class imbalance in the training dataset. Overall, the developed model achieved an accuracy of 80.00% in classifying all quality grades of the ‘Perfect Gold 111’ tomato variety, as shown in Table 6.

**Figure 5** Example images of tomato quality classification generated by the developed model: (A) Extra Class, (B) Class I, (C) Class II, and (D) Class III.

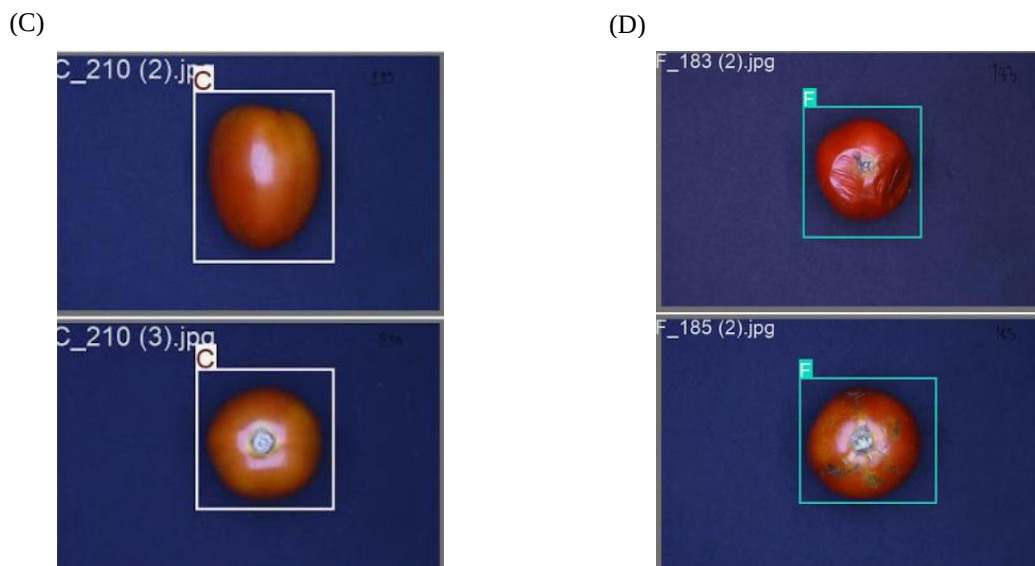


Figure 5 Example images of tomato quality classification generated by the developed model: (A) Extra Class, (B) Class I, (C) Class II, and (D) Class III. (Cont.)

Table 5 Confusion matrix of tomato quality classification for the ‘Perfect Gold 111’ variety generated using the developed model.

Predicted class (by developed model)	Extra Class	Class I	Class II	Class III	Total
Extra Class	11	4	0	0	15
Class I	3	26	2	0	31
Class II	0	0	5	1	6
Class III	0	1	0	2	3
Total	14	31	7	3	55

Table 6 Performance evaluation for tomato quality classification generated by the developed model

Quality class	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Extra Class	73.00	79.00	75.66	
Class I	84.00	84.00	83.92	
Class II	83.00	71.00	77.00	
Class III	66.00	66.00	65.91	
Overall				80.00

A comparative analysis of the results in Tables 3 and 4 reveals that the manual classification conducted by the farmers achieved an overall accuracy of 52.72%. This result indicates a relatively low level of classification accuracy, particularly for Class II and Class III, where Precision and Recall values were both below 60%. These findings suggest that farmers tended to misclassify tomatoes among classes with similar physical characteristics, such as size and skin color [16-18]. In contrast, Class I exhibited the highest Precision and Recall values among all categories (65% and 48%, respectively), indicating that farmers were more capable of accurately identifying tomatoes in this class than in the other classes [17, 19-25]. However, the overall Recall values remained low, reflecting a high tendency toward misclassification. This may be attributable to differences in subjective evaluation criteria, personal experience, and physical fatigue during manual grading [28].

A closer examination of the confusion matrix presented in Table 3 reveals that farmer misclassifications were not uniformly distributed across all class pairs, but were concentrated along specific class boundaries that reflect the inherent visual ambiguity of the ‘Perfect Gold 111’ variety. The most prevalent misclassification pattern involved the systematic confusion between Extra Class and Class I, as well as between Class I and Class II. This pattern may be attributed to three primary visual traits of this variety that are particularly challenging to assess through unaided human inspection.

First, the greenback area around the calyx exhibits a gradual, continuous color gradient rather than a distinct boundary, making it difficult for human graders to determine whether the extent of green discoloration falls within the threshold separating Extra Class from Class I under the TACFS 1503–2007 standard. Since the standard defines the Extra Class as permitting only a small characteristic greenback, whereas Class I allows minimal greenback that does not impair external appearance, the boundary between these two classes becomes inherently subjective during unaided visual inspection [17, 18].

Second, the length of shallow and linear scars on the tomato surface — key discriminators between the Class I (scar length ≤ 1 cm; linear scar $\leq$ two-thirds of tomato diameter) and the Class II — is extremely difficult to estimate accurately through visual inspection, particularly when scars are partially healed or when the tomato surface texture is uneven. Human graders tend to rely on a holistic visual assessment rather than precise linear measurement, which introduces systematic error at the Class I–Class II boundary [18].

Third, the suberization area at the calyx end — a corky, healed tissue whose extent defines the boundary between the Class I (≤ 1 cm²) and the Class II (≤ 2 cm²) — is another trait prone to human misestimation. Since suberization area requires two-dimensional spatial estimation, which humans perform poorly without calibrated measurement tools, this feature constitutes a major source of inter-grader inconsistency [18, 26].

In contrast, as shown in the confusion matrix in Table 5, the AI model demonstrated substantially reduced misclassification rates at the same class boundaries. This improvement is attributable to the model's capacity to perform pixel-level quantitative analysis of color distribution, defect geometry, and surface texture — effectively replicating the measurement-based criteria defined in the TACFS 1503–2007 standard at sub-millimeter precision. Specifically, the model's ability to analyze the spatial extent of greenback regions, the pixel-measured length of scars, and suberization area through image segmentation and feature extraction enabled classification boundary decisions to be based on objective visual evidence rather than subjective human estimation [18, 27, 30]. This mechanistic advantage plausibly explains the model's superior performance over farmer grading, particularly at the Extra Class–Class I and Class I–Class II boundaries, where human visual discrimination is inherently limited.

Conversely, the results presented in Tables 5 and 6 demonstrate that the developed model achieved an overall accuracy of 80%, indicating a substantially higher performance relative to farmer-based grading. Across all quality classes, the Precision, Recall, and F1-score ranged between 66% and 84%, reflecting a high and consistent level of classification performance. In particular, the Extra Class and Class I exhibited closely aligned Precision and Recall values, suggesting that the model was highly effective in distinguishing these quality levels. Although Class II and Class III exhibited slightly lower F1-scores (ranging between 65% and 77%), the model performance remained significantly higher than that achieved through farmer-based manual classification [23–25].

To statistically verify whether the observed difference in classification performance between the AI model (80.00%) and farmer grading (52.72%) was significant rather than attributable to chance, a chi-square test of independence was performed to compare the proportions of correct and incorrect classifications produced by the two graders across the same 55 test tomatoes. The observed frequencies used in the analysis are summarized in Table 7.

Table 7 Observed frequencies of correct and incorrect classifications generated by the AI model and farmers across 55 test tomatoes.

Classifier	Correct	Incorrect	Total
AI Model	44	11	55
Farmers	29	26	55
Total	73	37	110

The chi-square test yielded $\chi^2(1) = 9.18$, $p = 0.0024$, which is below the significance threshold of $\alpha = 0.05$. Therefore, the null hypothesis — that the AI model and farmer-based grading perform equivalently — was rejected. These results provide statistical evidence that the classification accuracy of the AI model was significantly superior to that of the farmers, and that the observed performance difference is not attributable to random variation.

The comparatively lower F1-Score for Class III (65.91%) may be attributed to the pronounced class imbalance in the training dataset. As shown in Table 2, Class I comprised 93 training samples compared to only 10 samples in Class III — a ratio of approximately 9:1. This underrepresentation likely limited the model's ability to learn discriminative features for Class III, resulting in a biased decision boundary toward the majority class and consequently reduced Precision and Recall for the minority class. To partially address this imbalance, data augmentation — a built-in function within the CiRA CORE platform — was applied during model training. However, this measure alone was insufficient to fully compensate for the substantial underrepresentation of Class III. Future studies should therefore incorporate targeted oversampling during data collection, synthetic minority oversampling techniques (SMOTE), or class-weighted loss functions during model training to improve classification performance for minority classes.

Collectively, these findings demonstrate that the deep learning model acquired robust representations of tomato quality characteristics — including color distribution, shape regularity, surface smoothness, and defect morphology — with a level of consistency and objectivity that manual inspection cannot reliably sustain [18, 26–27]. By training on a dataset encompassing diverse lighting conditions, viewpoints, and fruit sizes, the model developed generalized feature representations that reduce the perceptual variability and subjective bias inherent in human assessment [7, 27–29]. Critically, the model evaluates image-derived quality indicators — color

intensity, shape roundness, and defect extent — through pixel-level quantification rather than visual estimation, yielding repeatable decisions unaffected by grader fatigue or experiential inconsistency [28, 30].

4. Conclusion

This study developed an AI-based quality-grading model for the ‘Perfect Gold 111’ tomato variety using the CiRA CORE platform. The system included data collection, preparation, and model training. A comparative evaluation showed that farmers achieved an overall accuracy of 52.72%, while the developed model achieved an accuracy of 80.00%. The model outperformed human graders by learning quantitative image-based features, thereby reducing fatigue-related errors and inconsistency. These findings demonstrate the potential of artificial intelligence to enhance accuracy, efficiency, and consistency in agricultural grading and postharvest sorting processes. Nevertheless, several limitations of the present study should be acknowledged. First, the dataset comprised 220 tomato samples, which, while appropriate for a pilot investigation, represents a relatively limited sample size for deep learning applications. Second, model performance was evaluated using a single holdout train/test split rather than k-fold cross-validation; consequently, the reported accuracy of 80.00% may not fully reflect the model's generalizability across independent datasets. Third, the pronounced class imbalance in the training dataset — particularly the underrepresentation of Class III (10 samples) relative to Class I (93 samples) — contributed to the comparatively lower F1-Score for Class III (65.91%). Collectively, these limitations underscore the need for future work to expand data collection across multiple growing seasons and diverse field conditions, adopt stratified sampling strategies to achieve a more balanced class distribution, implement class-weighted loss functions or synthetic minority oversampling techniques to address class imbalance, and incorporate k-fold cross-validation to yield statistically rigorous performance estimates. These improvements would support the development of a robust, deployment-ready automated grading system for smallholder tomato production.

5. Acknowledgements

The authors gratefully acknowledge the support provided by Rajamangala University of Technology Isan through a grant from the Office of the Science, Research and Innovation Promotion Board (TSRI) under project code FF68/KKC/005. The authors also thank the Si Chiang Mai District Agricultural Office, Nong Khai Province, and the Phan Phrao large-scale tomato farmer group, Si Chiang Mai District, Nong Khai Province, for their valuable cooperation, provision of facilities, and supply of tomato samples used in this study.

6. Author Contributions

Doungpueng, K. : Conceptualization, Funding acquisition, Methodology, Writing – original draft; Pituprompan, K. : Methodology, Investigation, Resources; Lomchangkum, C. : Investigation, Resources; Premprayoon, K. : Investigation, Resources; Singhpoo, T. : Investigation, Resources, Software; Pachanawan, A. : Supervision, Project administration, Writing – review & editing

7. References

- [1] Liu W, Liu K, Chen D, Zhang Z, Li B, El-Mogy MM, Tian S, Chen T. *Solanum lycopersicum*, a model plant for the studies in developmental biology, stress biology and food science. *Foods* (Basel, Switzerland). 2022;11(16):2402.
- [2] Yosying S, Choenkwan S, Phumchantuk P, Palinthorn F. Fresh tomato production of small-scale farmers along Mekong River in northeastern region of Thailand. *KKU Research Journal (Graduate Studies)*. 2022;22(2):61-69.
- [3] Malawal P, Ueasin N. The study of technical efficiency in tomato production: A case study of Mekong River Bank, Thailand. *GMSARN Int J*. 2020;14(3):125-131.
- [4] Singh R, Jaiswal UC. An innovative machine learning framework for precise quantification of leaf disease severity via proposed hybrid segmentation and color manipulation technique. *Engineering Journal*. 2024;29(5).
- [5] Uddin MS, Bansal JC, editors. *Computer vision and machine learning in agriculture*. Singapore: Springer; 2021.
- [6] Khan A, Hassan T, Shafay M, Fahmy I, Werghi N, Seneviratne L, Hussain I. Tomato maturity recognition with convolutional transformers. *Sci Rep*. 2023;13:22885.
- [7] Wang S, Xiang J, Chen D, Zhang C. A method for detecting tomato maturity based on deep learning. *Appl Sci*. 2024;14(23):11111.

- [8] He Y, Fan B, Sun L, Fan X, Zhang J, Li Y, Suo X. Rapid appearance quality of rice based on machine vision and convolutional neural network research on automatic detection system. *Front Plant Sci.* 2023;14:1190591.
- [9] Nguyen TH, Nguyen HL, Bui NT, Bui TH, Vu VB, Duong HN, Hoang HH. Vision-based system for black rubber roller surface inspection. *Appl Sci.* 2023;13(15):8999.
- [10] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors.* 2018;18(8):2674.
- [11] Akther J, Nayan A-A, Harun-Or-Roshid M. Potato leaves blight disease recognition and categorization using deep learning. *Engineering Journal.* 2023;27(9).
- [12] National Bureau of Agricultural Commodity and Food Standards. Tomato (TACFS 1503–2007). Bangkok: Ministry of Agriculture and Cooperatives. 2007.
- [13] Noyunsan C, Katanyukul T, Saikaew K. Performance evaluation of supervised learning algorithms with various training data sizes and missing attributes. *Eng Appl Sci Res.* 2018;45(3):221-229.
- [14] Butkhot T, Reungsang P. Assessment of machine learning on sugarcane classification using Landsat-8 and Sentinel-2 satellite imagery. *Asia-Pac J Sci Technol.* 2021;26(04):APST-26-04-08.
- [15] Pornpanomchai C. Image-based orange (*Citrus reticulata* and *Citrus reticulata* Blanco) sweetness measurement by ResNet50. *Asia-Pac J Sci Technol.* 2025;30(05):APST-30-05-02.
- [16] Lu T, Han B, Chen L, et al. A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Sci Rep.* 2021;11:15824.
- [17] Vazquez DV, Spetale FE, Nankar AN, Grozeva S, Rodríguez GR. Machine learning-based tomato fruit shape classification system. *Plants.* 2024;13(17):2357.
- [18] Alzahrani M. Automated tomato defect detection using CNN feature fusion for enhanced classification. *Processes.* 2025;13(1):115.
- [19] Hou L, Liu Z, You J, Liu Y, Xiang J, Zhou J, Pan Y. Tomato sorting system based on machine vision. *Electronics.* 2024;13(11):2114.
- [20] Mputu HS, Abdel-Mawgood A, Shimada A, Sayed MS. Tomato quality classification based on transfer learning feature extraction and machine learning algorithm classifiers. *IEEE Access.* 2024;12:8283-8295.
- [21] Vo HT, Mui KC, Nguyen Thien N, Pham Tien P. Automating tomato ripeness classification and counting with YOLOv9. *Int J Adv Comput Sci Appl.* 2024;15(4).
- [22] Inuwa AB, Ahmad AS, Auwal ST, Muhammad M, Rabiul AL, Babangida I. Advances in sensor-based and machine learning techniques for automated tomato sorting and grading: A review. *J Sustainable Eng & Technol.* 2025;2(1):1-12.
- [23] Logeshwaran J, Srivastava D, Kumar KS, et al. Improving crop production using an agro-deep learning framework in precision agriculture. *BMC Bioinformatics.* 2024;25:341.
- [24] Hossen MI, Awrangjeb M, Pan S, et al. Transfer learning in agriculture: a review. *Artif Intell Rev.* 2025;58:97.
- [25] Sriprateep K, Khonjun S, Golinska-Dawson P, Pitakaso R, Luesak P, Srichok T, Chiaranai S, Gonwirat S, Buakum B. Automated classification of agricultural species through parallel artificial multiple intelligence system–ensemble deep learning. *Mathematics.* 2024;12(2):351.
- [26] Zuo H. Analysis and detection of tomatoes quality using machine learning algorithm and image processing. *Int J Adv Comput Sci Appl.* 2022;13(12).
- [27] Apostolopoulos ID, Tzani M, Aznaouridis SI. A general machine learning model for assessing fruit quality using deep image features. *AI.* 2023;4(4):812-830.
- [28] Zhou Z, Zahid U, Majeed Y, Nisha, Mustafa S, Sajjad MM, Butt HD, Fu L. Advancement in artificial intelligence for on-farm fruit sorting and transportation. *Front Plant Sci.* 2023;14:1082860.
- [29] Chuquimarca LE, Vintimilla BX, Velastin SA. A review of external quality inspection for fruit grading using CNN models. *Artif Intell Agric.* 2024;14:1-20.
- [30] Naqvi LH, Balasubramaniam B, Li J, Liu L, Li B. Four-dimensional hyperspectral imaging for fruit and vegetable grading. *Agriculture.* 2025;15(15):1702.