BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Deep Learning-Based Fruit and Vegetable Quality Grading Using Computer Vision: Recent Advances, Challenges, and Future Directions:Review

[1]Shubham Bhimrao Pote, [2]Prof. Jayshree Suryawanshi

[1]Email: poteshubham26@gmail.com
[2]Email: . jayshreesuryawanshi @gmail.com
[1,2]Department of Artificial Intelligence and Data Science,
[1,2]Ajeenkya DY Patil School of Engineering, Pune, Maharashtra India

Abstract

The grading of fruits and vegetables is mostly dependent on subjective visual inspection techniques that require labor. Computers and advanced machine learning systems can help to invent non-destructive tools for quality evaluation of fruits and vegetables based on color, size, shape, texture, ripeness, freshness, and surface damage based on computer vision methods. The paper is presenting reviews of methods used for fruit and vegetable grading, including convolutional neural networks, transfer learning, object detection systems, vision transformers, hyperspectral and multimodal vision techniques, and few-shot learning approaches. The review focuses on grading effectiveness, computation complexity, data variety, and algorithms reliability, and it shows that although many studies report the accuracy level higher than 99% while working with particular datasets, most studies used private datasets, making the comparison of results between various studies impossible. In this regard, the paper addresses the issue of a gap between benchmarks and real-life deployment and proposes a framework based on the grading process that includes a confidence lock mechanism for uncertain predictions and a reliable protocol for testing and benchmarking performances of the algorithms This paper lacks any experimental data as the proposed framework and procedures are designed for future assessment.

Keywords: Computer vision, deep learning, edge AI, explainable AI, fruit quality classification, hyperspectral imaging, transfer learning, vision transformer

I. INTRODUCTION

The purpose of quality checks in the fresh-produce supply chain is to ensure that produce is fresh, tasty, and safe. Grading is still done by trained inspectors, which requires time, skill, and subjectivity. Computer vision provides a non-destructive way of measuring visible characteristics that includes colour, size, shape, texture, maturity, and coating problems.

Deep learning is primarily replacing the previously used hand-crafted descriptors. A preliminary review of 40 studies in agriculture indicates that deep learning is generally more successful than the traditional image-

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

processing methods. A more recent review presents deep-learning machinery vision applied to the assessment of the quality of fruits and vegetables, and gives various examples of classification, detection, and segmentation of the fruits and vegetables images, maturity assessment, non-destructive evaluatio.

The reported accuracy is soaring on various datasets nowadays: from 97.96 to 99.50% across several fruit types [8], 99% average for classification and grading on one dataset with six fruit types [9], and from 97 to 100% accuracy for good versus rotten classification using a vision transformer [10]. However, a systematic review of 56 scientific papers on fruit inspection showed that most studies used their datasets, and therefore a comparison of models wasn't possible due to different datasets and collection conditions [2]. Thus, high benchmark accuracy doesn't necessarily mean that grading system will work properly when lighting, camera, fruit variety, or background changes.

The paper discusses deep-learning grading with an emphasis on robustness, efficiency, and application, instead of accuracy. The contributions are: (1) a technical background and a systematic comparison of recent reviews and selected primary studies; (2) an evidence-backed assertion about the benchmark-to-deployment gap; and (3) a grading framework focused on deployment and a testing methodology to be validated in the future. Section VII gives answers to the research questions, while Sections VIII–X provide limitations, future work, and conclusion. The paper describes no new experiments.

II. SCOPE, RESEARCH QUESTIONS AND METHOD

A. Scope and research questions

The paper discusses grading fruits and vegetables using their image data. Grading includes grade level, ripeness, freshness and visible defects on the fruits and vegetables. Research question framework includes five research questions that allow for thorough data analysis:

- Research Question 1: What are distinctions present in quality tasks when using CNN, the philosophy of transfer learning, detection and transformer architectures?
- Research Question 2: What are the effects of diversity in datasets and the conditions for acquiring datasets on generalization?
- Research Question 3: How is a trade-off created between accuracy, robustness, latency, memory, cost, and explainability?
- Research Question 4: When do multi-view, hyper-spectral, or multi-modal inputs yield advantages over RGB input?
- Research Question 5: What protocol meets the requirement of demonstrating deployment robustness rather than performance in benchmark testing?

B. Review method

This article provides a narrative overview of the processes of deep learning fruit and vegetable quality evaluation based on the functionalities of computer vision. The relevant scientific works were discovered through Google Scholar, IEEE Xplorer, and some other related online sources of scientific literature. The key topics of the search were fruit quality evaluation, deep learning, and modern methods of quality evaluation. The review process involved the analytical assessment of approximately 21-50 selected works.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Some articles were analyzed in detail while others were studied basing on their summaries. The articles have been chosen on the basis of their relations to the themes covered in the text, for instance, the models of deep learning, the methods of quality evaluation, the criteria of evaluation, the problems of working with datasets, and other issues. It is important to mention that the article presents a narrative review of deep learning use and that there is no attempt to perform a systematic review which means that there is no methodological protocol created for this review. During the evaluation process, some quantitative results were obtained, they were credited to original works, and were thoroughly analyzed knowing that methods of evaluation and experimental details differ a lot across.

III. TECHNICAL BACKGROUND

A. Acquisition and preprocessing

The use of RGB cameras takes up most of the studies involving fruit inspection (accounting for 84% of studies) [2]. LED lighting is usually found in industrial environments, whereas in 61% of studies images were captured using ambient light [2]. Normal preprocessing steps include resizing, normalisation, filtering, taking background out and data augmentation [2]. Given that the preprocessing processes may also have an influence on the performance of the model being developed, it is important to have all preprocessing steps documented, and data augmentation applied only to training data.

B. CNNs and transfer learning

Convolutional neural networks work on the images' inputs and are the default choice for classification tasks. Pre-trained networks can be adapted to grading purposes by transfer learning [11]; the most popular networks in the fruit-inspection literature include ResNet, VGG, Inception, and DenseNet [2]. The use of residual connections allows for training very deep networks, which is difficult otherwise [12]. EfficientNet enables scaling of the network depth, width, and input resolution at a fixed ratio [13] and has been applied for fruit grading [9] and, with attention mechanisms, for grading vegetables [14]. Transfer learning does not necessarily require pre-training on natural images. Aldakhil and Almutairi have trained a fruit-classification neural network first and transferred the model's parameters to a second model that performs quality grading; they refer to this as same-domain transfer learning [9].

C. Detection and segmentation

The distinction between segmentation and detection is that segmentation distinguishes between background and items whereas detection identifies positions of objects or defects in the image with bounding boxes and class labels. For instance, YOLO is an example of a single-stage detector which is performing the detection through a single-regression step over the images which is good for environments where latency is a concern. For example, improved YOLOv10 citrus defects detector trained on 1,270 original images (7,620 images actually used after augmentation) covering healthy oranges and four types of diseases showed in practice 97.7% mAP@50 value. The authors give their own assessment of the limitations: only four types of diseases without enough quantity of images obtained in reality with differentiation in lighting and overlapping lesions.

D. Attention and vision transformers

Vision transformers segment images into image patches which are then processed by self-attention [17]. They can interpret the global context but, unlike CNNs, they usually require a large-scale pre-training, on

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

account of no assumptions being built-in [17]. One ViT-based model has already been trained over a number of fruit datasets to better classify fruit into good or rotten [10].

E. Few-shot and vision–language learning

Few-shot and meta-learning methods are aimed at problems related to limited labeling. When it comes to fine fruit quality classification, ordinal meta-learning managed to achieve 85.12% accuracy, trailing behind the score of 77.23% obtained with standard meta-learning and the meager 21.49% garnered by few-shot transfer learning [18]. Vision-language models yield joint image-text representations that help implement zero-shot recognition [19]; a model that uses multiple characteristics for the purpose of language-image classification was proposed to identify fruit quality [20].

F. Spectral and multi-view sensing

Hyperspectral imaging conveys certain spectral information apart from the RGB. When coupled with the deep learning it was reviewed with respect to quality control and inspection of different products and their safety [7]. The data on the use of non-destructive imaging and spectroscopy in quality control is summarized [6]. The sensor fusion of deep learning in multimodal fruit evaluation has also been reviewed [21]. However, the images taken from a single point of view can capture only part of the subject whereas a multi-view acquisition would require additional equipment and calibration.

G. Explainability and confidence

Grad-CAM generates a map of the image regions that play a pivotal role in making a class prediction [22]. Although the maps can reveal whether the model is focusing on the object of interest rather than on the background, they are qualitative and, by themselves, are not sufficient to prove that an output is correct or false. In addition, modern networks are badly calibrated, which translates into unreliable confidence scores [23].

H. Evaluation metrics

In classification performance we present accuracy measures, including precision (P), recall (R) and F1-score, as well as confusion matrix for visualizing class-wise errors. In particular,

$$P = TP / (TP + FP), \quad R = TP / (TP + FN)$$

$$F1 = (2PR) / (P + R)$$

where TP, FP and FN stands for true positives, false positives and false negatives. Localisation performance is represented by intersection over union (IoU) metric that compares area under predicted region A and its reference region B; mean average precision (mAP) measure takes into account all classes and measures area under precision–recall curve averaged over all classes at specified IoU value:

$$IoU = |A \cap B| / |A \cup B|, \quad mAP = (1/K) \Sigma APk \qquad (3)$$

Continuous features, such as amount of sugar found in fruits or fruit colour score, could be presented by mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination R^2 metric:

$$MAE = (1/n) \Sigma |yi - \hat{y}i|$$

$$RMSE = \sqrt{[(1/n) \Sigma (yi - \hat{y}i)^2]}$$

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

$$R^2 = 1 - \Sigma (yi - \hat{y}i)^2 / \Sigma (yi - \bar{y})^2$$

The variable yi represents the actual values, $\hat{y}i$ refers to the predicted values, and $\bar{y}$ indicates the average actual value. Since the metrics are calculated on different datasets, label sets, and protocols, we do not rank the studies according to the reported figures.

TABLE I: QUALITATIVE SUMMARY OF TECHNIQUES FOR PRODUCE GRADING

Technique	Role in grading	Evidence and caveats
CNN with transfer learning [11]	Baseline classifier or grader	ResNet and VGG are common choices [2]; results depend on labelled data and capture conditions.
ResNet [12]	Deep backbone	Residual connections ease optimisation of deep networks.
EfficientNet [13]	Accuracy–size trade-off backbone	Applied to fruit [9] and vegetable [14] grading on single-source datasets.
YOLO [15]	Localisation and defect detection	Single-pass detector; example with stated GPU-only evidence [16].
Vision transformer [17]	Global-context backbone	Benefits from large-scale pre-training; applied to binary quality labels [10].
Few-shot / meta-learning [18]	Scarce-label settings	Accuracy depends strongly on the learning regime.
Vision–language models [19]	Zero-shot recognition	Proposed for fruit quality [20]; few grading studies in this review.
Grad-CAM [22]	Visual explanation	Qualitative; checks attention, not correctness.

IV. LITERATURE REVIEW

Table II contrasts fourteen recent reviews along with governing primary articles. Rojas Santelices et al. set three application scenarios, including orchards, food processing sectors, and marketplaces or end-use points, each with their own various set of requirements [2]. The figures stated may be considered as they appear in literature and cannot be treated as comparable through rows.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

TABLE II: COMPARISON OF REVIEWED STUDIES

Ref. / year	Study type and reported finding	Limitation or relevance
[1] 2023	*Survey, fruit ripeness classification.* Machine and deep learning dominate top-performing methods; deep learning avoids crop-specific engineered features.	Ripeness only; no single technique suits all fruit types.
[2] 2025	*Systematic review (PRISMA), 56 fruit-inspection studies, 2015–2024.* RGB in 84% of applications; 74% own datasets, 24% benchmark datasets; ResNet and VGG with transfer learning common.	Open-access journal articles only; fruit only; no direct comparison across studies.
[3] 2018	*Survey of 40 deep-learning studies in agriculture.* Deep learning generally outperformed commonly used image-processing techniques.	Broad scope, not grading-specific; predates vision transformers.
[4] 2025	*Review, deep-learning machine vision for produce quality.* Covers progress, challenges and prospects.	Chinese-language article; full text not assessed here.
[5] 2026	*Review, deep-learning fruit and vegetable image analysis.* Covers classification, detection, segmentation, datasets, metrics and trends.	Image analysis broadly; not grading-specific.
[6] 2024	*Review, external quality via non-destructive sensing.* Covers grading criteria, imaging and spectroscopic methods, preprocessing and reported accuracies.	External quality only; sensing-oriented, not deployment-oriented.
[7] 2025	*Review, hyperspectral imaging with deep learning.* Covers safety, internal and external quality; reports advantages in accuracy and efficiency.	Hyperspectral only; practicality in production still needs improvement.
[8] 2024	*Primary study: FruitVision grading model.* 97.96–99.50% accuracy across nine fruit categories (5-fold cross-validation).	Cross-validation within datasets; cross-condition or deployment results not stated in abstract.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[9] 2024	*Primary study: EfficientNetV2, same-domain transfer learning.* About 99% average test accuracy for classification and grading on six fruits (FruitNet); augmentation for class imbalance.	Single benchmark dataset; deployment measures not stated in abstract.
[10] 2023	*Primary study: general ViT model, multiple fruit datasets.* 97–100% accuracy for good-versus-rotten classification on most fruits.	Binary quality labels; no graded scale.
[14] 2024	*Primary study: EfficientNet with coordinate and CBAM attention, vegetables.* About 94–95% accuracy on six vegetables with three grades each (3,600 supermarket images; 480 training and 120 test images per vegetable).	Single source and one fixed split; headline accuracy differs between two tables of the paper (95.12% and 94.23%).
[16] 2025	*Primary study: YOLOv10-LGDA, citrus defects.* 97.7% mAP@50 on 1,270 original images (7,620 after augmentation), healthy fruit and four diseases.	Four diseases; few real-orchard lighting and occlusion images; GPU only (authors' statement).
[18] 2025	*Primary study: ordinal meta-learning, few-shot fruit quality.* 85.12% few-shot accuracy vs 77.23% (meta-learning alone) and 21.49% (few-shot transfer learning).	Few-shot regime; not directly comparable with fully supervised studies.
[24] 2026	*Preprint: apple instance segmentation, image-to-video shift.* Frame-level accuracy does not guarantee stable object identity; image-based metrics overestimate video performance.	Apples only; preprint, not peer reviewed.

Three trends can be identified. Firstly, there is consensus among studies that a move has taken place from hand-engineered features to deep learning [3], [1], [2], which have been extended from RGB classification to non-destructive and spectral sensing [6], [7]. Secondly, most papers dealing with CNN, EfficientNet, ViT and YOLO specific approaches have reported their accuracy varying from approximately 94% to 99% on their datasets [8], [9], [10], [14], [16]. Thirdly, label-scarce systems such as few-shot learning achieve lower accuracy in return for the effort required to annotate data with labels [18]. The limitations observed within these studies refer to the scope of datasets, labels and testing conditions rather than the model architecture itself.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

V. CRITICAL ANALYSIS AND RESEARCH GAP

The accuracy of benchmarks is remarkably high and frequently comes close to following a trend of convergence. A number of research works indicate 99% as the extent of an achieved accuracy [8], [9], [10]; however, 94 to 95% and 97.7% mAP@50 is also reported [14] and [16]. At this point the difference in performance of various architectures is not telling about a model's outcome in a new environment.

Difficulties with comparing studies are pointed out by Rojas Santelices et al., who state that 74% of the studies utilize proprietary datasets, while 24% rely on benchmark sets indicating that it is impossible to compare models properly since the datasets, conditions and targets are various [2]. They also mention the absence of agreements concerning the best configurations and propose to establish some uniform standards [2].

This information is shown in Fig. 1.

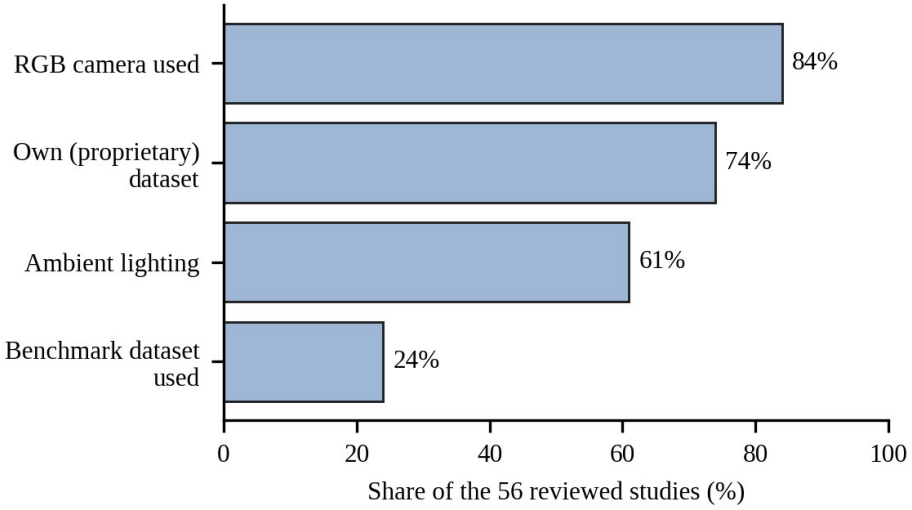


Figure 1. Features of the 56 studies pertaining to fruit evaluation as reviewed in [2]. The percentage is defined as in the review [2] (categories can exist together); drawn by the writer.

Conditions may differ among studies, but most are carried out under static conditions. Indeed, most studies are conducted in a lab and with static capture [2], which poses a question of varying light and occlusion. As noted by Wang et al., their detector does not contain enough real orchard images with varied lighting and overlapping lesions [16]. Test sets can be quite small, as shown by a study given in [14], where tests for vegetable evaluation of 120 images were performed using one source..

Image-level metrics can overstate real performance. A An article that analysed segmentation of apples, presented in a preprint format, indicated that models based on still images and rated by conventional metrics, appear to show varying forms of objects as well as counting mistakes in cases where light and reflections change in the video [24]. Although this is an example of commercially available material, it is still in the pre-peer review phase, which proves that having a high benchmark score does not ensure a model's proper performance.

Generalization to different agricultural products is challenging. Ripeness classification studies demonstrate that it is hardly possible to reach a universal solution to any type of fruit [1], while few-shot accuracy

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

depends considerably on the paradigm employed to learn [18]. Advanced sensors help to intensify the information but, according to experts, need to be more practical for agricultural production [7].

Research in the field of machine learning supports this conclusion. It has been determined that deep learning networks can take advantage of irrelevant signals such as background information [25], and it is common for modern networks to be improperly calibrated [23]. Even though they refer to general knowledge and do not involve any research on fruit grading, studies should be carried out to discover whether these findings are relevant to the grading process.

Research gap. Point established: deep models score very high accurately on the dataset they are intended. Point not clarified: articles published seldom test models outside their initial working conditions or present performance and confidence parameters beside their accuracy, while research done cannot be compared as different datasets and procedures are applied. Importance: a grader that works well in laboratory conditions might be prone to fail without notice under new lighting, cultivar, or camera, thus misgrading the produce during its harvest, manufacturing, or selling. Contribution of the paper: methodological frameworks and protocols presented in section VI aimed to reveal and address the problem.

VI. PROPOSED FRAMEWORK AND EVALUATION PROTOCOL

The framework and protocol below are this review's proposals. They have not been implemented or validated.

A. Deployment-oriented grading framework

The proposed pipeline is outlined in Fig. 2. As the basic strategy, RGB imaging is chosen; however, hyperspectral img, multi-view img and other sensing options can serve as branches. There are several baseline models: Compact CNN for baseline model, ResNet [12] and EfficientNet [13]; YOLO [15] is used as a localizing algorithm and vision transformer [17] or CNN–transformer type model can be used as advanced model. The few-shot method or vision–language technique [19] can be considered as possible additional ans when effective number of labels is low. Then, the estimated confidence is accompanied by a saliency explanation [22] that is sent to the confidence gate in order to separate the untrusted data from the trusted data.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

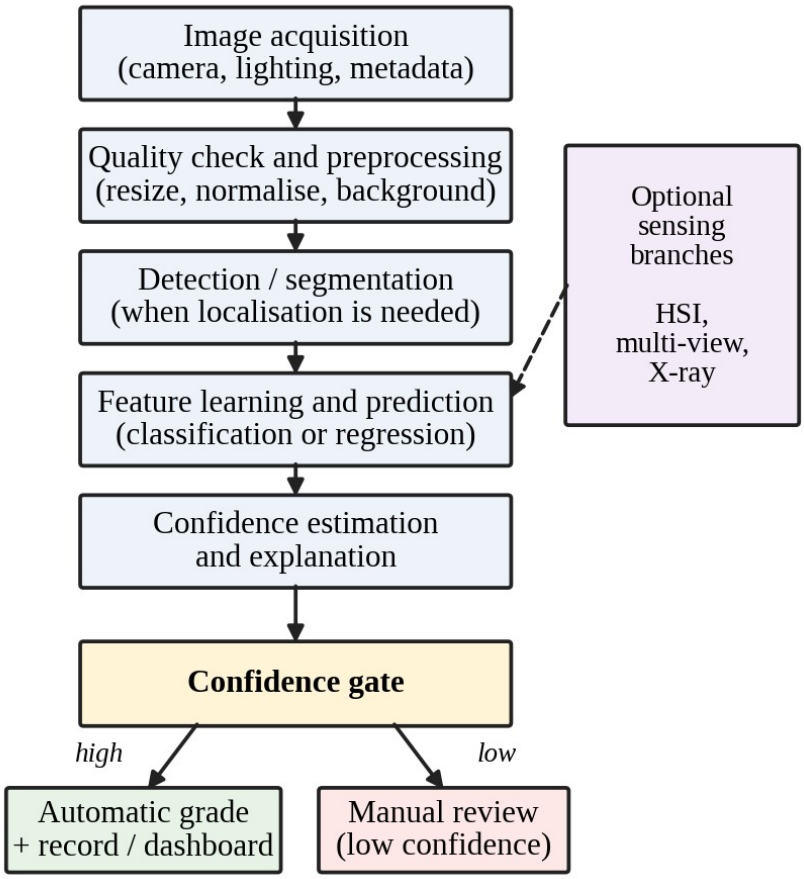


Fig. 2. Proposed deployment-oriented grading framework with a confidence gate and optional sensing branches (original diagram).

B. Evaluation protocol

Fig. 3 describes the protocol, Table III contains the evaluation parameters. The dataset, as well as the commodity, definition of labels, acquisition procedure, distribution of classes and license, has to be indicated. It is unnecessary to separate different views of the same real fruit, repeating frames as well as samples of one batch or session into training and test datasets due to information leakage via near-duplicates. Additionally to in-distribution tracking data, the test set collected in various conditions should be withheld as well. It is also imperative to inform about the training configuration, used random seed, optimizer, learning rate schedule, stopping criterion, hardware and version of the model. The metrics follow the expressions given in (1)-(6) and reported per class and per condition as totals alone do not show errors in minority classes or neighboring grades. The deployment report contains the information about inference time, throughput., model size and peak memory on the target hardware.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

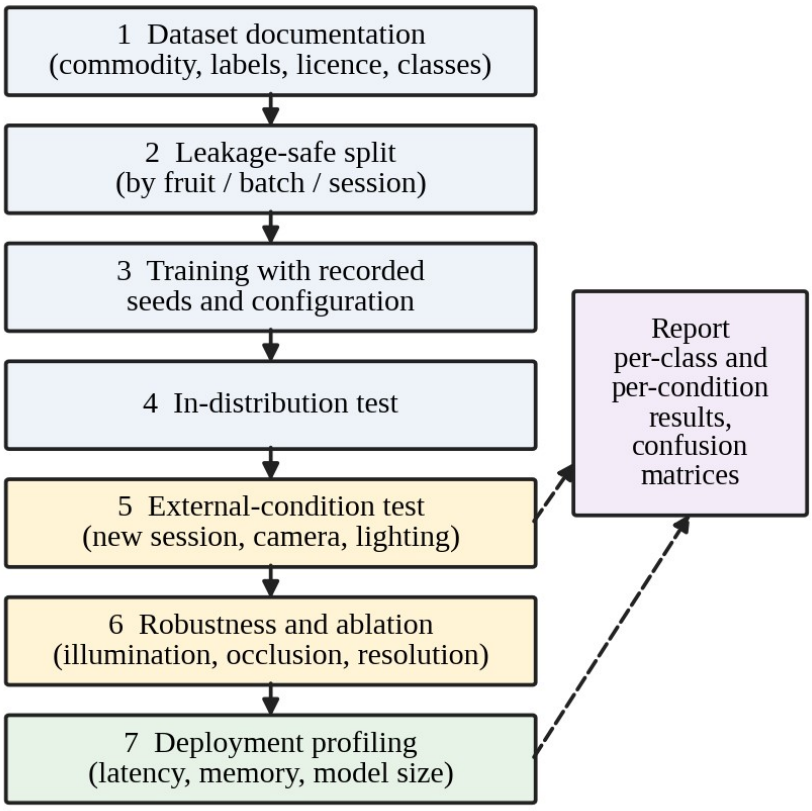


Fig. 3. Proposed reproducible evaluation protocol (original diagram).

TABLE III: PROPOSED EVALUATION DIMENSIONS

Dimension	Measures	Purpose
Predictive performance	Accuracy, macro-F1, precision, recall, confusion matrix; mAP/IoU; MAE, RMSE, R^2	Class-wise results expose errors hidden by aggregate accuracy.
Cross-condition robustness	Results on an external-condition test set; breakdown by illumination, background, viewpoint and occlusion	Tests dependence on acquisition conditions [25].
Confidence and calibration	Expected calibration error; accuracy versus rejection rate	Supports routing uncertain samples to manual review [23].

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Efficiency	Latency, throughput, model size, peak memory, hardware	Establishes deployment feasibility.
Explainability	Saliency maps on correct and incorrect cases [22]	Checks that the model attends to the produce, not the background.
Reproducibility	Seeds, optimiser, schedule, stopping rule, hardware, model version, data licence	Allows others to repeat the study.

Planned ablations cover lighting and colour-temperature changes, occlusion, background clutter, viewpoint, input resolution, augmentation, model complexity against latency, and RGB-only versus multimodal input.

VII. DISCUSSION

RQ1 (architecturesResNet and VGG are standard models in fruit inspection and are widely used, according to the research [2]. The performance metrics of EfficientNet-based [9], [14], ViT-based [10], and YOLO-based [16] approaches show an accuracy between 94% and 99% (or 97.7% mAP@50). The YOLO approach performs the functions of defect detection and localisation while the other methods are used in the classification of images. The main reason for the inability of literature to rank architectures is that different datasets, splits and conditions are used, and therefore the models need to be compared directly according to one benchmark**.**

RQ2 (databases alongside conditions). 74% of research-related papers mentioned the use of proprietary datasets [2], as well as having shown that there is no universal method that would work properly on all the types of fruit per se. Therefore, the results tend not to be generalized due to the fact that primary authors themselves acknowledge that it is crucial to have a different lighting during experiments [16]. Work done in other databases and under various conditions has vital importance for following the protocols.

RQ3 (tradeoffs). The sources that were analyzed within this section do not provide a systematic framework for calculating tradeoffs involved in the process of achieving a good balance between accuracy, reliability, speed, responsiveness, costs and explainability, since accuracy and results differ significantly from source to source while even the best-performing detectors provide results measured on the basis of GPU [2]; one more fact to prove the idea is that the best detector provides results only in terms of its performance on machine learning platform [16]. Profiling and calibration processes presented in table III help to provide the knowledge that is absent.

RQ4 (modalities). By means of using hyperspectral imaging along with deep learning technology it is possible to evaluate external

RQ5 (Protocol). Section 6 suggests a protocol consisting of the following: leakage-safe splitting, external-condition testing, robustness ablations, calibration and deployment profiling. However, the usefulness of the protocol has not been verified yet.

Use cases. Orchards, processing lines and retailer settings have different technical requirements. While confidence gate is applicable in high-throughput lines with manual review lane, scenarios with different

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

lighting conditions such as retail and orchards place much more emphasis on robustness tests. This rationale is the author's viewpoint rather than the conclusion of the above-mentioned publications.

VIII. LIMITATIONS AND THREATS TO VALIDITY

The purpose of this narrative review is not to undertake a systematic review, however, it should be noted that the selection of studies was biased toward those that presented high accuracy, which resulted in 14 publications being listed in table 2. Hence, only abstracts, reviews or summaries of the primary studies were used, instead of original sources. Additionally, it should be outlined that one study is in Chinese language, and one study is a preprint. Moreover, it should be stressed that the figures have appeared in different experiments/settings and therefore are not subject to comparison. What is more, the framework described is yet underdeveloped, and no results are available concerning performance, latency and durability of such framework. While the quality grades appear to be subjective, it should be reaffirmed that RGB images cannot measure certain internal characteristics; because of this the gap between benchmark conditions and those applied in real time is an inferred one and needs further measurement.

IX. FUTURE RESEARCH DIRECTIONS

This methodology could also have been applied to a specially designed multi-condition dataset to determine the degree of success of competitors as well as to consider various additional research directions, such as adaptation across the domains, self-supervised learning, and active learning with limited annotation, model compression through knowledge distillation (26), quantization (27), and other technologies for compressed model deployment into the edge devices, improvement of algorithms to combine RGB and hyperspectral data, creating multi-view frameworks to produce 3D models of produce samples, uncertainty evaluation, and methods for the integration of IoT or robotic systems for sorting.

X. CONCLUSION

With deep learning, grading of agricultural products has shifted from relying on feature engineering to using a variety of techniques, such as CNNs (especially EfficientNets), detectors, vision transformers, and few-shot methods, with accuracy on individual datasets exceeding 94% and reaching as high as 99% (see Table II). Nevertheless, the scientific literature provides little information concerning how efficient particular models are across different datasets and areas of application or under different conditions associated with their utilization. The development of a framework and methodology for testing vision-based models for grading of produce is the main contribution of the present study.

ACKNOWLEDGMENT

AI assistance statement: The authors of this study utilized AI-powered tools, such as generative artificial intelligence (AI), to assist them in proofreading and editing their article during the peer-review stage. The authors assert that they are responsible for the correctness, authenticity and innovative content of the written paper and declare that they did not employ any AI tool for creating the main ideas of the manuscript. This statement should comply with the AI policy established by the corresponding journal when the manuscript was accepted for publication.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

REFERENCES

[1] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," *Artif. Intell. Agric.*, vol. 7, pp. 44–57, 2023, doi: 10.1016/j.aiia.2023.02.004.

[2] I. Rojas Santelices, S. Cano, F. Moreira, and Á. Peña Fritz, "Artificial vision systems for fruit inspection and classification: Systematic literature review," *Sensors*, vol. 25, no. 5, Art. no. 1524, 2025, doi: 10.3390/s25051524.

[3] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[4] Y. Yan, Y. Yu, T. Kong, F. He, and Z. Li, "Deep learning-based machine vision for intelligent perception of fruit and vegetable quality: Progress, challenges, and prospects," *Food Sci.*, vol. 46, no. 22, pp. 23–39, 2025 (in Chinese), doi: 10.7506/spkx1002-6630-20250401-007.

[5] J. Ma, Y. Wan, W. Min, Z. Ma, L. Tan, and S. Jiang, "Frontiers and advances of deep learning-based fruit and vegetable image analysis," *Comput. Electron. Agric.*, vol. 241, Art. no. 111256, 2026, doi: 10.1016/j.compag.2025.111256.

[6] T. Akter *et al.*, "A comprehensive review of external quality measurements of fruits and vegetables using nondestructive sensing technologies," *J. Agric. Food Res.*, vol. 15, Art. no. 101068, 2024, doi: 10.1016/j.jafr.2024.101068.

[7] C. Yang *et al.*, "Hyperspectral imaging and deep learning for quality and safety inspection of fruits and vegetables: A review," *J. Agric. Food Chem.*, vol. 73, no. 17, pp. 10019–10035, 2025, doi: 10.1021/acs.jafc.4c11492.

[8] A. Hayat, F. Morgado-Dias, T. Choudhury, T. P. Singh, and K. Kotecha, "FruitVision: A deep learning based automatic fruit grading system," *Open Agric.*, vol. 9, no. 1, Art. no. 20220276, 2024, doi: 10.1515/opag-2022-0276.

[9] L. A. Aldakhil and A. A. Almutairi, "Multi-fruit classification and grading using a same-domain transfer learning approach," *IEEE Access*, vol. 12, pp. 44960–44971, 2024, doi: 10.1109/ACCESS.2024.3379276.

[10] I. D. Apostolopoulos, M. Tzani, and S. I. Aznaouridis, "A general machine learning model for assessing fruit quality using deep image features," *AI*, vol. 4, no. 4, pp. 812–830, 2023, doi: 10.3390/ai4040041.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019, vol. 97, pp. 6105–6114.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[14] J. Wen and J. He, "Agricultural development driven by the digital economy: Improved EfficientNet vegetable quality grading," *Front. Sustain. Food Syst.*, vol. 8, Art. no. 1310042, 2024, doi: 10.3389/fsufs.2024.1310042.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[16] L. Wang, R. Ye, Y. Chen, and T. Li, "YOLOv10-LGDA: An improved algorithm for defect detection in citrus fruits across diverse backgrounds," *Plants*, vol. 14, no. 13, Art. no. 1990, 2025, doi: 10.3390/plants14131990.

[17] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021, arXiv:2010.11929.

[18] A. Mishra, M. Patwardhan, P. Deshpande, and B. Rai, "Ordinal-meta learning for fine-grained fruit quality prediction," in *Computer Vision – ECCV 2024 Workshops* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2025, pp. 303–318, doi: 10.1007/978-3-031-91835-3_20.

[19] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. 38th Int. Conf. Mach. Learn. (ICML)*, 2021, vol. 139, pp. 8748–8763.

[20] J.-L. Duan, L.-Q. Lai, Z. Yang, Z.-J. Luo, and H.-T. Yuan, "Multi-feature language-image model for fruit quality image classification," *Comput. Electron. Agric.*, vol. 227, Art. no. 109462, 2024, doi: 10.1016/j.compag.2024.109462.

[21] R. Singh, R. Nisha, R. Naik, K. Upendar, C. Nickhil, and S. C. Deka, "Sensor fusion techniques in deep learning for multimodal fruit and vegetable quality assessment: A comprehensive review," *J. Food Meas. Charact.*, vol. 18, no. 9, pp. 8088–8109, 2024, doi: 10.1007/s11694-024-02789-z.

[22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[23] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, 2017, vol. 70, pp. 1321–1330.

[24] K. Park, A. Pal, and J. H. Mok, "Understanding Image2Video domain shift in food segmentation: An instance-level analysis on apples," 2026, arXiv:2602.08491 (preprint, not peer reviewed).

[25] R. Geirhos *et al.*, "Shortcut learning in deep neural networks," *Nat. Mach. Intell.*, vol. 2, no. 11, pp. 665–673, 2020, doi: 10.1038/s42256-020-00257-z.