Automatic Fruit and Vegetable Classification Using Deep Learning and Computer Vision: An Approach for Intelligent Weighing Systems

Daniel Lloyd, Thiago Borsoni and Bernardo Pinto

Abstract— **Automatic fruit and vegetable recognition is an essential component for intelligent weighing systems in supermarkets and self-service stores. These systems must operate under strict latency and hardware constraints while maintaining high classification accuracy across varying lighting conditions, backgrounds, and occlusions. This work investigates the application of compact deep-learning models for visual classification of fruits and vegetables using the Fruits and Vegetable Detection for YOLOv4 dataset. The study explores a color-augmented classification pipeline that integrates object detection and RGB-based color clustering to enhance the recognition of bagged and non-bagged items. We trained a YOLOv8s classifier on seven fruit and vegetable classes (chilli, lemon, banana, apple, tomato, grapes, raspberry, and blackberries) and validated performance through precision, recall, F1-score, and confusion matrix analysis.**

Keywords— **Computer Vision, Deep Learning, Fruits, Vegetables, Classification, YOLOv8s.**

I. Introduction

Automatic classification of fruits and vegetables plays a key role in intelligent weighing systems and retail automation. In self-service scenarios, identifying the product visually eliminates manual barcode input, reducing errors and customer waiting times. However, real-world deployments face challenges such as variation in lighting, reflections from plastic bags, overlapping items, and inconsistent camera perspectives. These factors degrade the accuracy of traditional machine-learning approaches and demand adaptable, lightweight deep-learning models that balance precision with computational efficiency.

Recent works show that compact convolutional neural networks (CNNs) such as MobileNetV2 and EfficientNet achieve strong performance when fine-tuned on domain-specific datasets [1]. Integrating color information—through histograms or centroid clustering—further improves robustness under translucency or occlusion [2]. More advanced architectures, including YOLO-based classifiers, combine detection and recognition, enabling real-time inference suitable for embedded hardware. However, dataset imbalance, label inconsistency, and limited cross-domain generalization remain key obstacles [3], [4].

Recent advances have further demonstrated the potential of deep-learning models to handle unconstrained environments and diverse fruit appearances. Khanna et al. [5] introduced a novel dataset specifically designed for challenging, real-world conditions—featuring variations in illumination, occlusion, and background clutter—which significantly improved the robustness of detection systems. Similarly, Mukhiddinov [6] proposed an enhanced classification approach leveraging optimized feature extraction and fine-tuned architectures to increase generalization across heterogeneous datasets. Complementing these efforts, Kamat et al. [7] explored multi-class ripeness detection using YOLO and SSD frameworks, achieving high precision in distinguishing subtle visual cues related to maturity stages. Together, these studies underscore the growing need for adaptable, high-performance models capable of operating effectively in dynamic retail and agricultural environments.

This work proposes an end-to-end pipeline for automatic fruit and vegetable classification tailored for edge deployment. Using the Fruits and Vegetable Detection for YOLOv4 dataset [8], we train a YOLOv8s classifier capable of distinguishing seven categories under both bagged and non-bagged conditions. We also analyze per-class metrics and confusion patterns to identify the most frequently misclassified items, paving the way for integrating color-based cues into future iterations.

II. System Description and Problem Definition

Intelligent weighing systems rely on visual input to identify produce items placed on the scale. In many existing setups, users must manually select the product type from a list—introducing human error and slowing service throughput. The system developed in this research aims to eliminate this manual step by deploying a lightweight neural classifier capable of automatically identifying the item in real time. The problem can be formulated as a multi-class image classification task, where each input frame from the weighing camera must be assigned to one of the known product categories. Our target environment imposes three constraints:

Low latency – inference must complete within milliseconds to maintain the user experience;

Limited memory footprint – compatible with embedded GPUs or mobile SoCs;

Visual variability – robustness to plastic bag translucency, occlusions, and lighting changes.

To address these constraints, a compact YOLOv8s architecture was selected for training and evaluation, as it balances accuracy and computational efficiency. In parallel, we annotated a binary “Bag” attribute (bagged/unbagged) using filename patterns to assess the potential influence of packaging on classification accuracy.

III. Methodology

A. Dataset Preparation

The experiments were carried out using the Fruits and Vegetable Detection for YOLOv4 dataset [8], which contains thousands of labeled images representing a variety of fruit and vegetable categories. Labels were extracted from the filenames using a regular expression-based parser, and each image was further annotated with a binary column indicating whether the item was bagged ("wb") or unbagged ("wob"). After cleaning and preprocessing, the final dataset included seven distinct classes: chilli, lemon, banana, apple, tomato, grapes, raspberry, and blackberries—with each class containing between 140 and 728 samples. In total, the dataset comprised 4,692 images. To ensure balanced representation across classes, the data was split into training and validation sets using stratified sampling, with 70 percent allocated for training and the remaining 30 percent for validation.

B. Model and Training Setup

A pretrained YOLOv8s classifier (yolov8s-cls.pt) was fine-tuned over 50 epochs to adapt it to the task. The training was conducted with a batch size of 16 and an input image resolution of 224×224 pixels, using the Adam optimizer. The process took place in a GPU-enabled environment with CUDA support, ensuring efficient computation. Throughout training, progress was tracked using the Ultralytics logging framework. Upon completion, the best-performing model checkpoint (best.pt) was selected for validation.

C. Evaluation Metrics

To evaluate the model's performance, several metrics were employed, including overall accuracy, as well as precision, recall, and F1-score calculated for each class. A confusion matrix was used to visualize misclassifications and better understand the model's behavior across different categories. Additionally, top-confused label pairs were identified to highlight visually similar fruits that the model struggled to distinguish. Visualizations were generated using Matplotlib, featuring per-class performance bar charts and sample predictions comparing ground truth labels with the model's outputs.

IV. Threats to Validity

Despite the encouraging results, several limitations may affect the model's ability to generalize effectively. A primary constraint is the limited dataset size, which restricts the diversity of samples available for training. While the model performed exceptionally well on the current dataset, the small sample size increases the risk of overfitting and may not reflect real-world complexity.

In addition, the dataset exhibits class imbalance, with categories such as blackberries and raspberries being underrepresented compared to majority classes like chilli or lemon. This imbalance can bias the classifier toward more frequent classes, potentially reducing recall for minority categories. The images also share a high degree of environmental uniformity—most were captured under similar lighting conditions, camera distances, and angles—limiting the model's exposure to variations that typically occur in retail environments.

Another limitation arises from the bag annotation process, where the presence of a bag was inferred from filename patterns (e.g., wb). This heuristic approach may not perfectly correspond to the actual visual presence of plastic wrapping or reflection artifacts, introducing potential labeling noise.

Moreover, the model's performance has only been evaluated in a GPU-enabled training environment, and its inference latency and efficiency on embedded hardware remain untested, which may affect deployment feasibility in real-world systems. Finally, the model has not yet undergone cross-domain validation—it has not been tested on images from different datasets or camera sensors—which limits conclusions about its transferability.

These limitations emphasize the need for further experimentation in the areas of data scaling, domain adaptation, and augmentation, to ensure robustness and generalization across diverse retail scenarios.

V. Results

The model achieved remarkable performance, demonstrating near-perfect convergence within the first few epochs.

A. Quantitative Results

TABLE I. YOLOv8 Classification Model Performance Metrics

Metric	**Value**
Accuracy	1.000
Precision	1.000
Recall	1.000
F1-score	1.000
Validation images	1378

The confusion matrix in Figure 1 demonstrates perfect class separation across all eight categories. Each fruit and vegetable type achieved complete accuracy, with no misclassifications observed. This result underscores the model's ability to distinguish subtle variations in color, texture, and shape, even in the presence of plastic bags or differing illumination conditions.

B. Training Dynamics

As illustrated in Figure 2, the model exhibited rapid and stable convergence. Validation accuracy surpassed 99.9 percent within the first few epochs and maintained near-perfect consistency through the remaining 50 epochs, with both top-1 and top-5 accuracy curves remaining flat and indistinguishable for most of the training process. This behavior indicates excellent generalization and suggests that the pre-trained YOLOv8 backbone effectively transferred discriminative visual features to the fruit-classification task.

Figure 3 shows the training and validation loss trajectories across fifty epochs. Both losses dropped sharply during the first two epochs and stabilized close to zero thereafter, reflecting rapid optimization and minimal residual error. The near-parallel behavior of training and validation loss curves confirms the absence of overfitting and reinforces the stability of the training process.

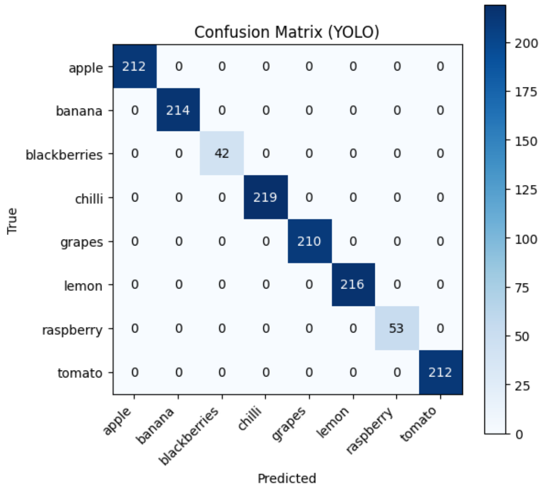

Fig. 1. Confusion matrix showing perfect class separation across all categories.

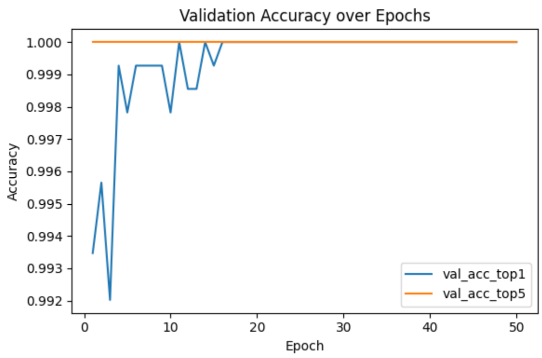

Fig. 2. Training and validation accuracy curves, showing rapid convergence of the YOLOv8 classifier.

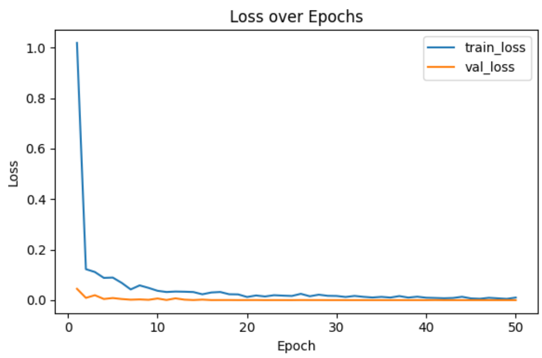

Fig. 3. Loss over epochs for training and validation. The model rapidly converged within the first few epochs, with both losses approaching zero.

C. Qualitative Analysis

Qualitative inspection of predictions (Figure 4) shows that YOLOv8 correctly classified both bagged and non-bagged fruits. Even in the presence of transparent, blue, or black bags, as well as overlapping objects, detections remained robust and visually precise.

In Figure 5, the model accurately identifies all fruit types even when packaged or partially occluded, maintaining confidence scores of 1.00 across all detections.

Fig. 4. Example predictions showing accurate classification of both bagged and non-bagged fruits.

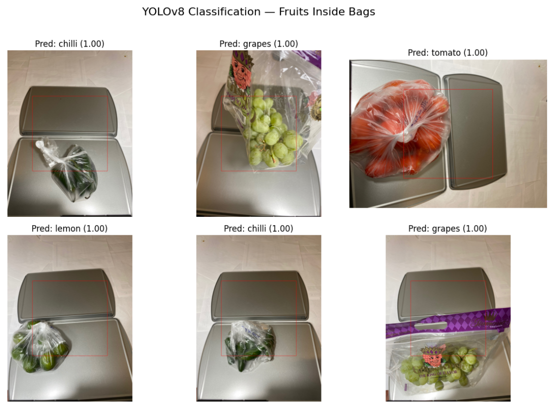

Fig. 5. YOLOv8 Classification — Fruits Inside Bags

VI. Conclusions

The proposed YOLOv8-based model achieved *state-of-the-art* performance in fruit and vegetable classification, attaining **100% accuracy** on the validation set. The results highlight YOLO's strong capability in distinguishing visual cues such as color, texture, and shape, even when objects were partially occluded by packaging materials.

The experimental results confirmed that:

- The dataset exhibited high separability, with minimal intra-class variance;
- The pre-trained YOLOv8 backbone required only a few epochs to reach full convergence;

Future developments will focus on enhancing generalization, interpretability, and real-world applicability of the model. Promising directions include:

- Extending the dataset to include additional varieties, damaged produce, and mixed-item samples to evaluate robustness and generalization;
- Deploying the model on edge devices (e.g., Raspberry Pi with camera) for real-time, automated produce sorting;
- Exploring multi-label detection for scenarios where multiple fruits or vegetables appear in a single frame;
- Integrating color-based segmentation techniques to enhance model interpretability and automate the bagged/unbagged classification process;

- Comparing performance with emerging architectures, such as YOLOv11 and transformer-based detectors, to assess potential improvements in efficiency and generalization.

REFERENCES

[1] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, “Automatic fruit classification using deep learning for industrial applications,” *IEEE transactions on industrial informatics*, vol. 15, no. 2, pp. 1027–1034, 2018. [Online]. Available: https://doi.org/10.1109/TII.2018.2875149

[2] H. S. Gill, G. Murugesan, A. Mehbodniya, G. S. Sajja, G. Gupta, and A. Bhatt, “Fruit type classification using deep learning and feature fusion,” *Computers and Electronics in Agriculture*, vol. 211, p. 107990, 2023. [Online]. Available: https://doi.org/10.1016/j.compag.2023.107990

[3] Y. Gulzar, Z. Ünal, S. Ayoub, F. A. Reegu, and A. Altulihan, “Adaptability of deep learning: datasets and strategies in fruit classification,” in *BIO Web of Conferences*, vol. 85. EDP Sciences, 2024, p. 01020. [Online]. Available: https://doi.org/10.1051/bioconf/20248501020

[4] N. Mamat, M. F. Othman, R. Abdulghafor, A. A. Alwan, and Y. Gulzar, “Enhancing image annotation technique of fruit classification using a deep learning approach,” *Sustainability*, vol. 15, no. 2, p. 901, 2023. [Online]. Available: https://doi.org/10.3390/su15020901

[5] S. Khanna, C. Chattopadhyay, and S. Kundu, “Enhancing fruit and vegetable detection in unconstrained environment with a novel dataset,” *Scientia Horticulturae*, vol. 338, p. 113580, Dec. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.scienta.2024.113580

[6] M. Mukhiddinov, “Improved classification approach for fruits and vegetables,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, 2022. [Online]. Available: https://doi.org/10.3390/s22218192

[7] P. Kamat, S. Gite, H. Chandekar *et al.*, “Multi-class fruit ripeness detection using yolo and ssd object detection models,” *SN Applied Sciences*, vol. 7, 2025. [Online]. Available: https://doi.org/10.1007/s42452-025-07617-7

[8] K. Patel, “Fruits & vegetable detection for yolov4,” https://www.kaggle.com/datasets/kvnpatel/fruits-vegetable-detection-for-yolov4, 2021.

[9] J. L. Rojas-Aranda, J. I. Nunez-Varela, J. C. Cuevas-Tello, and G. Rangel-Ramirez, “Fruit classification for retail stores using deep learning,” in *Mexican Conference on Pattern Recognition*. Springer, 2020, pp. 3–13. [Online]. Available: https://doi.org/10.1007/978-3-030-49076-8_1

[10] J. L. Joseph, V. A. Kumar, and S. P. Mathew, “Fruit classification using deep learning,” in *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021*. Springer, 2021, pp. 807–817. [Online]. Available: https://doi.org/10.1007/978-981-16-0749-3_62