COMP8430: Advanced Computer Vision and Action
Major Project Phase 1-1: Make a Proposal

Fine-Grained Fruit Classification: Challenges, Significance, and Dataset Collection Strategy

Student name: Phuong Thao (Jasmine) Huynh

Student ID: 46248722

Table of Contents

1. RESEARCH GAP AND MOTIVATION

Fine-grained fruit classification is a critical yet underexplored area in agricultural computer vision, with growing relevance in intelligent farming, post-harvest processing, and automated retail systems. While existing classification models can reliably differentiate broad fruit categories, such as apples, bananas, or grapes, they often fail to identify subtle distinctions between closely related cultivars or maturity levels, which are essential for quality grading, pricing, and inventory control (Rao et al., 2024). The need for accurate, cultivar-specific classification has intensified with the expansion of precision agriculture and the global fruit trade, where intra-class variation and inter-class similarity pose significant challenges for both manual inspection and automated detection systems (Afroj et al., 2024).

Despite advances in deep learning, most existing datasets are designed for coarse-grained classification tasks and collected under controlled conditions. These datasets often lack sufficient subclass diversity, exhibit class imbalance, and contain images captured under uniform lighting, limiting their generalizability to real-world agricultural environments (Yang et al., 2024). Furthermore, fruit classification models trained on such datasets tend to perform poorly in the presence of natural variability, including occlusion, uneven illumination, and cluttered backgrounds, which are common in orchards, markets, and household settings (Wu et al., 2024). These issues are further exacerbated when deploying models on edge devices, where limited computational resources constrain both model complexity and inference accuracy (Zhang et al., 2025).

To address these limitations, this study proposes a real-world, fine-grained fruit classification task focused on distinguishing between similar fruit cultivars and ripeness stages under diverse environmental conditions. A high-resolution image dataset will be developed using mobile phone photography and online sources, incorporating variations in lighting, viewpoint, and background context. This dataset will serve as a benchmark for training and evaluating lightweight yet discriminative deep learning models, with an emphasis on performance, robustness, and deployability in both agricultural and consumer-facing contexts. By bridging the gap between controlled research settings and practical field conditions, this work aims to advance the development of reliable and scalable fruit classification systems.

2. IMPORTANCE OF FINE-GRAINED FRUITS CLASSIFICATION

Fine-grained fruit classification is essential in modern agriculture, post-harvest processing, food retail, and intelligent automation. In contrast to broad fruit classification, which distinguishes fruits at the species level, fine-grained classification enables cultivar-level identification and ripeness assessment based on subtle visual cues such as color gradation, texture, and shape (Apostolopoulos et al., 2023). This level of precision supports high-value sorting, targeted handling, and variety-specific processing across the supply chain. The integration of deep learning has significantly improved the accuracy, speed, and scalability of classification systems, enabling real-time deployment in resource-constrained environments (Rao et al., 2024).

2.1 Enhancing post-harvest handling and supply chain efficiency

Accurate identification of fruit cultivars and maturity levels is critical for post-harvest grading, sorting, and packaging. Manual systems often result in inconsistencies, while automated fine-grained classification enables standardization and efficiency across industrial pipelines (Wu et al., 2024). In commercial contexts, subtle differences in fruit ripeness directly affect shelf life, pricing, and packaging strategies, making visual classification a key quality control step (Zhang et al., 2025). Deep learning models trained on diverse datasets can recognize such differences even under challenging conditions, supporting robust grading and operational scalability (Yang et al., 2024). These models also reduce reliance on human labor and support automated, high-throughput systems (Apostolopoulos et al., 2023).

2.2 Economic and commercial value in food retail and trade

In retail and wholesale markets, cultivar-level classification underpins traceability, pricing integrity, and fraud prevention. Many high-value cultivars—such as Alphonso mangoes or Honeycrisp apples, command different prices despite similar appearances, increasing the need for reliable automated systems (Afroj et al., 2024). AI-powered tools embedded in self-checkout stations and handheld scanners enable consistent classification, particularly for non barcoded or bulk produce (Rao et al., 2024). These systems also reduce human error, facilitate inventory tracking, and improve retail operations (Gill et al., 2022).

2.3 Applications in precision agriculture and harvesting robotics

Fine-grained classification supports intelligent agriculture by enabling optimal harvest timing and selective picking. Identifying ripeness stages improves yield quality and reduces waste, while cultivar-level recognition enables targeted harvest strategies

(Zhang et al., 2025). Deep learning models embedded in robotic harvesters assist in fruit localization even under partial occlusion or irregular lighting conditions, enhancing automation in dynamic environments (Wu et al., 2024). These models can also aid in early disease detection and anomaly recognition by identifying subtle surface changes (Yang et al., 2024).

2.4 Deployment in lightweight and embedded AI systems

The increasing demand for portable and field-deployable classification tools has led to the development of lightweight neural architectures. Models such as LightNN and MIRNet_ECA achieve competitive accuracy with reduced computational overhead, making them well-suited for use on smartphones, drones, and embedded devices (Afroj et al., 2024). Such systems enable real-time decision-making in remote or low-connectivity regions, particularly benefiting smallholder farmers and mobile inspection teams (Zhang et al., 2025). Edge-based classification also facilitates continuous monitoring and faster feedback cycles in agricultural operations (Rao et al., 2024).

2.5 Broader societal impact through smart applications and consumer tools

Fine-grained fruit classification underpins a range of consumer-facing applications, including smart dietary tools, nutrition recommendation systems, and digital kitchen assistants. These applications rely on real-time identification of fruit types to estimate nutritional content and detect allergens, supporting healthier and more informed food choices (Rao et al., 2024). Moreover, low-cost classification technologies empower small vendors to automate quality control and meet grading standards without access to specialized equipment (Gill et al., 2022). As digital agriculture expands, fine-grained classification plays a foundational role in ensuring inclusivity, transparency, and sustainability (Apostolopoulos et al., 2023).

3. CHALLENGES IN FINE-GRAINED FRUITS CLASSIFICATION

Despite advancements in deep learning and computer vision, fine-grained fruit classification remains a challenging and unresolved task. This is primarily due to the combined effects of intra-class variation, environmental inconsistency, model generalization limitations, and deployment constraints. In real-world agricultural scenarios, models must accurately differentiate between visually similar cultivars while operating under diverse conditions and resource limitations (Rao et al., 2024). Addressing these challenges requires innovation across dataset design, feature extraction techniques, and model deployment strategies (Zhang et al., 2025).

3.1 Visual similarity and intra-class variability

Fine-grained fruit classification is often hindered by the high morphological similarity between distinct cultivars and significant variability within the same class. For instance, cultivars such as Crown and Dangshan pears, or Crimson and Sweet Globe grapes, may share overlapping color, shape, and size characteristics, making visual differentiation difficult using conventional features (Rao et al., 2024). Additionally, natural variation in fruit size, surface texture, and ripeness caused by environmental factors introduces intra-class noise, reducing classification confidence (Afroj et al., 2024). These complexities increase the risk of misclassification, particularly when deploying models in uncontrolled environments (Gill et al., 2022).

3.2 Environmental interference in natural settings

Many classification models are trained on datasets collected under ideal or static conditions, limiting their effectiveness in dynamic orchard environments. Variability in lighting, occlusion from foliage, complex backgrounds, and inconsistent fruit orientation can degrade image quality and model performance (Zhang et al., 2025). In practical applications, such as classifying mango or dragon fruit ripeness, models trained on clean, indoor datasets often fail when exposed to cluttered field images with natural background noise (Wu et al., 2024).

3.3 Dataset imbalance and annotation complexity

The availability of balanced, high-quality training data remains a core limitation. Existing fruit datasets frequently suffer from class imbalance, where common cultivars dominate and rare or seasonal varieties are underrepresented, leading to biased learning outcomes (Rao et al., 2024). Furthermore, manual image annotation is labor-intensive and prone to inconsistency, particularly when differentiating between visually similar cultivars requires domain expertise (Apostolopoulos et al., 2023). Inaccurate or inconsistent labeling can reduce generalizability and limit reproducibility in experimental research (Gill et al., 2022).

3.4 Feature extraction and model generalization

Convolutional Neural Networks (CNNs) and vision transformers are effective for coarse-grained classification but often lack sensitivity to the fine-scale features needed to distinguish similar fruit types. CNNs tend to lose spatial detail in deeper layers, while attention-based models such as MSAPVT and MIRNet_ECA provide improved discrimination at the cost of computational complexity (Yang et al., 2024). Moreover, models pre-trained on general-purpose datasets frequently underperform in agricultural applications due to domain shift and insufficient fine-grained transferability (Afroj et al., 2024).

3.5 Deployment constraints on edge and mobile devices

Field-ready classification systems must operate on mobile or embedded devices with limited processing power, memory, and energy capacity. Balancing accuracy with efficiency is a persistent challenge, particularly for fine-grained classification where high-resolution visual features are essential (Wu et al., 2024). Lightweight models such as LightNN have been developed to reduce computational overhead while maintaining accuracy, yet their performance may degrade in complex conditions such as occlusion or lighting distortion (Zhang et al., 2025). Additionally, variability in mobile camera hardware and environmental conditions complicates real-time deployment and system robustness (Rao et al., 2024).

4. PROPOSED DATASET COLLECTION STRATEGY

4.1 Individual Work Within Group Project

Although this project is conducted in groups, both Phase 1-1 (proposal and dataset collection) and Phase 1-2 (model implementation) must be completed individually. Each student is required to independently define a classification task, collect and annotate an original dataset, and evaluate models accordingly. This proposal outlines a fine-grained fruit classification task focused on visually distinguishing between closely related fruit cultivars based on photographic and morphological characteristics.

4.2 Fine-Grained Class Selection and Image Sources

This project will target 21 distinct fruit cultivars selected from five widely consumed fruit types: **apples, oranges, grapes, bananas, and pears**. These subclasses were chosen based on their high commercial relevance, availability in both Australian and Southeast Asian markets, and their visual similarity, which poses practical challenges for real-world classification systems.

To ensure a robust and diverse dataset, **approximately 15 to 50 images** will be collected for each class, depending on cultivar availability and verification confidence. The expected total dataset size will range between **1,000 and 1,400 images**. This range reflects the project's emphasis on **label accuracy, environmental diversity, and high-quality visual data**, rather than uniform sample quantity.

Selected Fruit Subclasses:

Fruit Type	Cultivars (21 classes)
Apples (8)	Granny Smith, Pink Lady, Royal Gala, Fuji, Jazz, Kanzi, Modi, SnapDragon
Oranges (2)	Navel, Valencia
Grapes (5)	Thompson Seedless, Crimson Seedless, Autumn Royal, Cotton Candy, Sweet Globe
Bananas (2)	Cavendish, Lady Finger (Baby Banana)
Pears (4)	Beurré Bosc, Nashi, Corella, Red Angel

Data will be collected through a combination of manual photography and curated online sources:

- **Manual Photography:** Images will be captured using iPhone 11 and iPhone 12 Pro (Sydney), and iPhone X (Vietnam). Fruits will be photographed at supermarkets, wet markets, and domestic settings. Only pre-labeled fruits with clear signage, packaging, or tags will be selected to ensure reliable cultivar identification.

- **Online Sources:** Supplementary images will be sourced from Shutterstock, Alamy, and 123RF. Only images where the cultivar name is clearly provided in the image title or metadata will be included. Visual consistency with known cultivar morphology will also be a condition for inclusion.

No images will be taken from pre-existing benchmark datasets to comply with the assignment requirements.

4.3 Image Composition, Diversity, and Environmental Conditions

Images will be taken in real-world retail contexts (e.g., market displays, produce sections, home counters) and will not include fruit on trees or in natural orchards, in order to reflect commercial presentation settings. The focus will be on the fruit itself, photographed individually or in small groups, against light or neutral backgrounds where possible.

To enhance dataset variability and promote model generalizability, diversity will be introduced across the following dimensions:

Diversity Dimension	Implementation Plan

Color Variants	Collect different ripeness stages or blush patterns within each cultivar (e.g., red vs. yellow Fuji)
Backgrounds	Include retail shelf displays, wooden tables, light-colored packaging, home surfaces
Views and Angles	Top-down, side profile, close-ups, diagonal and partial crop shots
Lighting Conditions	Images under daylight, shade, artificial lighting, and minor low-light environments
Devices	iPhone 11, iPhone 12 Pro (Sydney); iPhone X (Vietnam)
Source Type	Mix of real-world mobile photography and verified online images
Image Format	All collected images will be saved in .jpg format to ensure compatibility across platforms. Any images originally in other formats (e.g., .png) will be converted to .jpg during preprocessing to maintain uniformity.

Future stages of the project (Phase 2) will incorporate additional variation by collecting images under rainy, dark, or flash-based conditions using alternative mobile devices, as required by the assignment brief.

4.4 Annotation Strategy and Label Accuracy Assurance

To ensure consistency and ease of integration with deep learning frameworks such as PyTorch and TensorFlow, all images will be stored using a standardized class-based folder structure. Each fruit cultivar will be assigned a dedicated subfolder, and all image files will be named using a uniform convention.

Folder Structure:

```
dataset/
  apple_fuji/
    apple_fuji_01.jpg
    apple_fuji_02.jpg
  grape_crimson_seedless/
    grape_crimson_seedless_01.jpg
```

grape_crimson_seedless_02.jpg

Image Naming Convention:

Each image will be named using the format:

[fruit type]*[cultivar name]*[index].jpg

This ensures clarity, simplifies class-level indexing, and supports reproducibility across experiments.

Examples:

- apple_granny_smith_01.jpg
- pear_red_angel_17.jpg
- grape_autumn_royal_04.jpg

The file extension for all images will be standardized to **.jpg** to maintain compatibility with common data preprocessing pipelines.

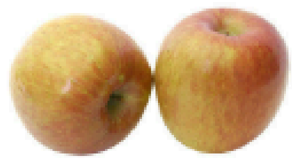
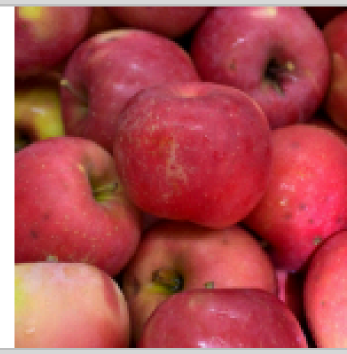
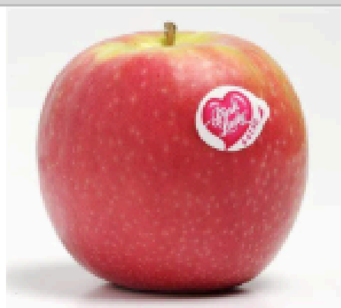
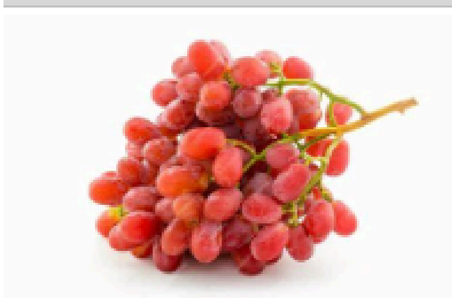
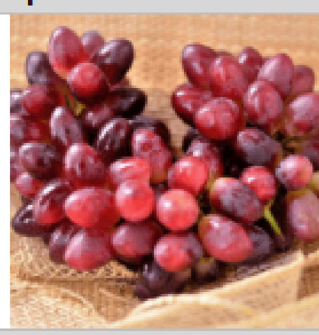
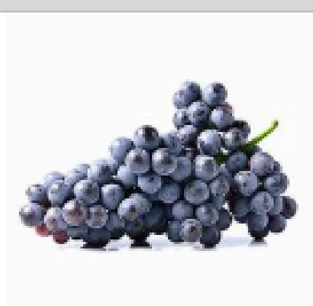
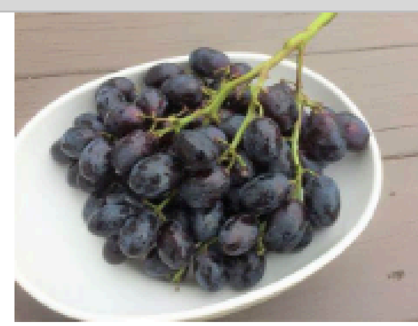
Label Accuracy Assurance:

Given the fine-grained nature of this task, label accuracy is paramount. To mitigate labeling errors and maximize dataset integrity, the following multi-step verification protocol will be applied:

1. Manual label capture at source: Only fruits with cultivar labels displayed clearly on packaging, price tags, or signage in markets and supermarkets will be photographed. This ensures that labels are based on verified commercial naming, not visual estimation.
2. Online image verification: Only images explicitly labeled with cultivar names in the title or metadata on platforms such as Shutterstock, Alamy, or 123RF will be included. Images will be screened manually to confirm they reflect known morphological characteristics of the cultivar.
3. Cross-check with these sources (if necessary):
    - Official cultivar databases from government agriculture departments and fruit producer associations (e.g., NSW DPI).
    - AI-powered plant identification tools, such as *PlantNet* or *PictureThis*.
    - Morphological comparisons against reference images and descriptors, including size, shape, blush pattern, stem type, and skin texture.

Images with ambiguous labeling or lacking visual clarity will be discarded to maintain high dataset quality.

4.5 Visual Examples from the Dataset:

Apple - Fuji		
		
Apple - Pink Lady		
		
Grape - Crimson Seedless		
		
Grape - Autumn Royal		
		

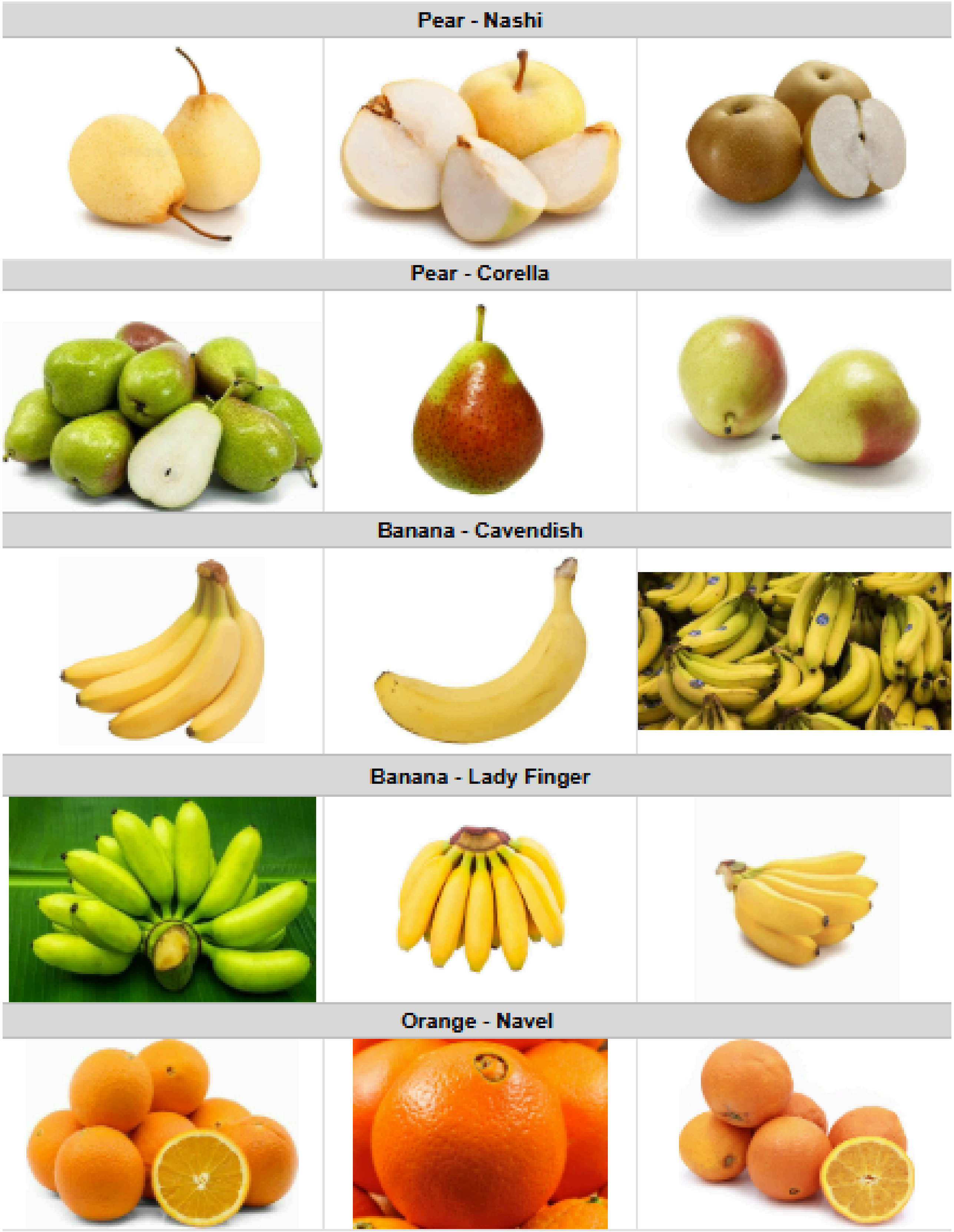


Figure 1. Example fruit images in the dataset collected from manual and online sources.

4.6 Timeline and Feasibility Plan

Date	Milestone
March 21–23	Finalize list of cultivars; submit Phase 1-1 draft for preview
March 24–30	Complete image collection (≈15–50 images per class); finalize raw dataset
March 30–31	Organize file structure and filenames; conduct label verification
April 1–3	Split dataset into training (60%), validation (10%), and test (30%) sets
April 4–6	Implement baseline models (e.g., ResNet50, InceptionV3); begin model experimentation
April 7–8	Train and evaluate models; generate visualizations of misclassified images
April 9–10	Analyze results; assess performance across categories; finalize visual outputs
April 11	Submit Phase 1-1 and Phase 1-2 reports, including annotated dataset summary and visual results

REFERENCES:

Afroj, M., Jahan, I., Islam, M. P., & Rahman, M. S. (2024, September). LightNN: An optimized deep learning framework for enhanced fruit classification. In *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)* (pp. 1–6). IEEE. https://doi.org/10.1109/COMPAS60761.2024.10796630

Apostolopoulos, I. D., Tzani, M., & Aznaouridis, S. I. (2023). A general machine learning model for assessing fruit quality using deep image features. *AI, 4*(4), 812–830. https://doi.org/10.3390/ai4040041

Gill, H. S., Khalaf, O. I., Alotaibi, Y., Alghamdi, S., & Alassery, F. (2022). Multi-model CNN-RNN-LSTM based fruit recognition and classification. *Intelligent Automation & Soft Computing, 34*(3), 1351–1365. https://doi.org/10.32604/iasc.2022.022589

Rao, Y., Li, C., Xu, F., & Guo, Y. (2024). MSAPVT: A multi-scale attention pyramid vision transformer network for large-scale fruit recognition. *Journal of Food Measurement and Characterization, 18*, 9233–9251. https://doi.org/10.1007/s11694-024-02874-3

Wu, T., Miao, Z., Huang, W., Han, W., Guo, Z., & Li, T. (2024). SGW-YOLOv8n: An improved YOLOv8n-based model for apple detection and segmentation in complex orchard environments. *Agriculture, 14*(11), 1958. https://doi.org/10.3390/agriculture14111958

Yang, X., Zhao, W., Wang, Y., Yan, W. Q., & Li, Y. (2024). Lightweight and efficient deep learning models for fruit detection in orchards. *Scientific Reports, 14*, 26086. https://doi.org/10.1038/s41598-024-76662-w

Zhang, B., Lou, K., Wang, Z., Xia, Y., Fu, W., & Bai, Z. (2025). MIRNet_ECA: Multi-scale inverted residual attention network used for classification of ripeness level for dragon fruit. *Expert Systems with Applications, 274*, 127019. https://doi.org/10.1016/j.eswa.2025.127019