**VeggieScan**

**Visual Diagnosis of Vegetable Freshness and Contamination**



**A Capstone Project Presented to the Faculty**

**of the College of Computer Studies**

**St. Michael’s College**

**Iligan City**

**By**

**Kerneil Rommel S. Gocotano**

**John Michael S. Abiol**

**June 2025**

**Chapter 1**

**PROBLEM AND ITS SETTING**

**Introduction**

According to the latest global assessment by the Food and Agriculture Organization (FAO), roughly one-third of all food produced for human consumption—around 1.3 billion tons per year is lost or wasted somewhere along the supply chain [1]. Fresh fruits and vegetables account for a significant portion of this loss, due to their high moisture content and rapid metabolic activity, which make them especially prone to spoilage. But food waste isn’t just an economic or environmental concern it’s also a direct public health risk. The World Health Organization (WHO) estimates that unsafe food causes 600 million cases of foodborne illness and 420,000 deaths annually, with fresh produce cited as one of the primary transmission vectors [2]. These impacts fall hardest on vulnerable populations, including children, the elderly, and communities without consistent access to refrigeration or cold storage.

A range of advanced tools have emerged in response to rising concerns over produce spoilage. Hyperspectral imaging systems are now used to detect pigment degradation and early-stage mold on leafy greens [3], while bio-impedance spectroscopy offers non-destructive estimates of cellular integrity and freshness [4]. Portable biosensors have also shown promise for rapid, on-site screening of chemical and microbial contaminants [5]. However, these technologies remain largely confined to research labs and industrial packing houses. Their cost, physical footprint, and reliance on specialized workflows make them impractical for use in the informal retail environments where most shoppers, especially in low- and middle-income settings, make day-to-day purchasing decisions.

Because professional instruments are often inaccessible, most consumers rely on simple visual cues like color, gloss, or signs of wilting to judge freshness. But studies using eye-tracking and hedonic ratings show just how subjective these judgments can be: even minor changes in hue or edge sharpness can lead to wildly different perceptions of quality [6]. As a result, shoppers are left asking “Is this still good?” with little more than gut instinct—leading to unnecessary waste, or worse, health risks. Recent advances in mobile hardware and cloud-based AI offer a potential alternative. Prototype smartphone apps can already identify product types and estimate ripeness from a single photo [7], while large-scale evaluations show that modern computer vision systems can now rival human inspectors in assessing color, size, and surface defects [8]. Still, most of these tools stop short of what matters most: delivering real-time, plain-language feedback that everyday shoppers can actually use.

VeggieScan is designed to bridge that gap. The system combines cross-platform image capture with server-side computer vision and a lightweight language model layer to deliver instant, easy-to-understand results: “Good to Buy,” “Buy with Caution,” or “Not Recommended.” In doing so, it brings lab-grade assessment to the pocket level empowering everyday consumers to make safer, more informed choices while reducing household food waste and minimizing the risk of foodborne illness.

**Statement of the Problem**

Iligan City, Lanao del Norte, is a rapidly growing urban center where open-air markets such as Pala-o, Tambo, and Tibanga remain the primary distribution hubs for leafy greens and other vegetables. Local climate data show daytime temperatures ranging from 29 °C to 34 °C year-round, with minimal relief at night [9]. In stalls lacking refrigeration or even electric fans produce is often piled onto wooden tables or plastic tarpaulins. This setup creates trapped heat, poor airflow, and compression damage, accelerating wilting, bruising, and microbial growth.

A 2024 Citizen Participatory Audit described “mountain-like” accumulations of spoiled vegetables at the Bonbonon materials-recovery facility and linked the problem directly to rapid stall-side spoilage and the absence of cold storage infrastructure [10]. A separate senior high school survey of night-market patrons reported common lapses in hygiene including bare-hand handling and limited washing stations and found that 71% of respondents feared foodborne illness when buying raw vegetables, yet had no reliable way to assess freshness on site [11].

Together, these local findings highlight three tightly connected concerns:

1. Accelerated spoilage: High ambient heat, direct sunlight, and poor ventilation quickly degrade produce quality, promote surface contamination, and shorten shelf life.
2. Subjective inspection: With no objective tools, vendors and buyers rely solely on visual guesswork—leading to premature discards, poor-quality purchases, and uncertainty about safety.
3. Cascading waste and risk: Spoiled produce feeds directly into the growing organic waste stream, while invisible contamination increases the chance of foodborne illness for households.

Addressing these challenges demands a low-cost, stall-side tool that can deliver real-time, objective feedback on freshness and visible contamination tailored to the realities of Iligan’s wet markets. Without such an intervention, the city is likely to see continued food loss, financial strain on both vendors and consumers, and avoidable health risks at the household level.

**Objectives of the Study**

This study aims to develop and evaluate *VeggieScan*—a smartphone-enabled visual diagnostic system tailored to the buying conditions of Iligan City’s public markets. Its overarching goal is to equip shoppers and vendors with instant, evidence-based assessments of vegetable freshness and safety, thereby reducing household food waste and minimizing foodborne risk.

Specifically, the study will:

1.Compile a representative image dataset of the most commonly sold vegetables from the Pala-o, Tambo, and Tibanga markets, covering distinct freshness stages.

2.Train and fine-tune a computer vision model, paired with a language-generation layer, to classify freshness levels and flag visible signs of spoilage or contamination.

3.Deploy the model through a cross-platform mobile and web interface that operates effectively under typical market lighting and connectivity conditions.

4.Evaluate system performance over a three-month field trial in Iligan, focusing on diagnostic accuracy, user trust, and the projected reduction in spoilage-related waste.

**Scope and Limitation of the Study**

This project will deliver *VeggieScan* as a two-part system meant to work within the actual conditions of Iligan City’s public markets. The first part is a mobile app, built in Flutter, that runs on Android 11 or iOS 15 and above. It lets users take one to three photos of a vegetable under normal stall lighting and submit them with a single tap. The second part is a web portal built with React.js, designed for use on laptops or kiosks, mainly for vendors who prefer larger screens. Both the mobile app and web portal send images to a Python Flask backend hosted on a cloud server physically located in the Philippines to keep latency low for local users. Once the server receives the image, it applies OpenCV preprocessing auto white balance, contrast normalization, and background masking—to clean and prepare the input before analysis.

The cleaned image is then passed to a fine-tuned visual–language model based on LLaMA-3, trained on a curated dataset of the twelve most common vegetables sold in the Pala-o, Tambo, and Tibanga markets. The model predicts three freshness levels and checks for five visible issues: mold, insect damage, abnormal liquid, bruising, and wilting. Those predictions go through a lightweight decision layer that turns them into a single verdict—“Good to Buy,” “Buy with Caution,” or “Not Recommended” along with a short, plain-language reason. During the three-month field pilot, all predictions and usage logs will be stored for analysis of accuracy and user trust, but no personally identifiable data or photos will be kept for more than 30 days.

VeggieScan only works with visual evidence it can’t detect internal rot, chemical residue, or anything not visible on the surface. Model accuracy is also limited to the conditions it was trained on: images collected in Iligan between July 2025 and January 2026. That means performance may drop for out-of-season imports or rare vegetable types not included in the dataset. The system assumes decent lighting; if the stall is too dark or the lighting is too color-biased, the model might misclassify freshness. It also needs a stable internet connection 4G or Wi-Fi for real-time feedback. Users in network dead zones may experience delays or upload failures. Although the backend can scale up, it’s currently limited to around 100 requests at once, so crowded market hours might still cause short wait times. Offline usage isn’t supported yet, since everything runs through the server.

Lastly, VeggieScan is only meant to assist, not decide for the user. It gives suggestions based on visual data, but the final choice still rests with the buyer. It doesn’t guarantee protection against foodborne illness or financial loss and shouldn’t be treated as a replacement for basic caution or common sense.

.

**Significance of the Study**

This part discusses the significance of developing VeggieScan, a mobile- and web-based visual diagnostic tool for vegetable freshness and contamination in Iligan City, Philippines. The researchers believe that the project will be beneficial to the following:

**Consumers.** This study will help consumers make quick, informed decisions about vegetable freshness and safety, reducing food-borne illness risk and minimizing spoilage losses.

**Vendors** This study will help vendors receive immediate product feedback, enabling them to adjust pricing, displays, and restocking to reduce waste and maintain customer trust.

**Current Researchers.** This study will help the researchers to enhance knowledge with the current research.

**Future Researchers.** This study will help future researchers to use the system as a reference.

**Definition of Terms**

**AI-Driven Recommendation Engine**: Module that converts model outputs into the consumer-facing verdicts: “Good to Buy,” “Buy with Caution,” or “Not Recommended.”

**Bruisin**g: Barkened or softened tissue caused by mechanical damage; treated here as a visible spoilage cue.

**Cloud Backend**: Python-based server hosted in a Philippine data center; handles image processing and model inference.

**Computer Vision** (CV): Automated image analysis using OpenCV and a trained classification model.

**Contamination Indicator**: Any visible sign suggesting safety risk, including mold, insect damage, exudate, bruising, or wilting.

**Dataset**: Collection of labeled vegetable images used for model training and testing.

**Freshness Score**: Numeric output (range: 0–1) indicating how closely an image matches the “fresh” class.

**Image Pre-processing**: Adjustments like color correction and contrast normalization applied before classification.

**OpenCV**: Open-source computer vision library used for image enhancement and tensor conversion.

**RGB Image**: Standard color photograph with red, green, and blue channels; used as the system’s only input format.

**Surface Mold**: Fungal growth appearing as fuzzy or discolored patches on the vegetable surface.

**Wilting**: Loss of leaf firmness due to moisture loss; a visible sign of aging or spoilage.

**Chapter 2**

**RELATED LITERATURE AND RELATED STUDIES**

This chapter deals with the review of related literature. The purpose of the study of research works done in the same field is to understand what type of study has been explore previous research projects in the same topic. The study related literature and research work is very essential as it provides the researcher proper guidelines. This chapter also includes a synthesis literature and studies that help the researchers used to gather ideas and information.

**Related Literature**

He et al. approached a long-standing issue in post-harvest handling: how to reliably grade leafy greens that degrade quickly once taken out of cold storage. Their team collected 8,200 hyperspectral image cubes from spinach and Chinese cabbage across a simulated four-day shelf life. To reduce computational load, they compressed the original 256 spectral bands into 16 principal components before feeding the data into a hybrid CNN-GRU architecture. The CNN was tasked with picking up on spatial micro-textures—early signs of wilting and surface pitting while the GRU tracked subtle temporal shifts in reflectance, particularly at 680 nm and 730 nm, wavelengths associated with chlorophyll degradation. Training was carried out over 12,000 iterations, using a step-decay learning rate of 0.015 and a fixed batch size of 32; image patches were resized to 128 × 128 pixels. When benchmarked against a human sensory panel, the model achieved a macro-F1 score of 97.6%, and more impressively, it flagged produce as “end-of-life” a full day ahead of expert graders. Saliency maps showed the model’s attention gravitating toward faint yellowing at the leaf margins subtle features that even experienced shoppers typically miss. It’s a good case study for how machine vision can catch quality signals hiding beneath the surface noise. [12].

In a separate study, Koyama and colleagues tested whether ordinary smartphones—rather than specialized spectrometers could capture enough visual information to assess freshness. Over the course of 30 days, they photographed spinach bundles under both fluorescent and LED lighting at daily intervals, collecting 2,700 RGB images. Green-channel dominance in the histograms ranged from 12% to 28%. From these, they extracted a mix of HSV values, GLCM textures, and Sobel-edge density features. These were fed into a radial-basis function SVM using a gamma of 0.1 and a cost parameter of 32. Despite relying on consumer-grade camera optics, the model still managed 84% accuracy across a three-tier freshness scale. Precision stayed above 0.80 even when lighting conditions were inconsistent. Notably, fusing color and texture data outperformed color alone by 11 percentage points—highlighting the importance of fine-grained features like leaf turgor. They also clocked mean inference times of 0.42 seconds on a mid-range Snapdragon 720G, suggesting that on-device prediction could be viable in high-traffic retail environments.[13]

The “GoMicro–Spotcheck” project took portability even further by swapping out bulky optical setups for a $1 clip-on microscope offering 20× magnification and built-in ring lighting. To test its viability, researchers captured just 50 images per class across okra, tomato, eggplant, bitter gourd, and chili. With such limited data, they leaned heavily on augmentation introducing ±20° rotations, brightness shifts between 0.9 and 1.2, and Gaussian noise with σ = 12—to expand each class to 2,000 samples.They fine-tuned a MobileNet-V3 Small model, set with a 0.35 depth multiplier and 224-pixel input, over 40 epochs using Adam (learning rate 0.001, β₁ = 0.9). Field trials at Adelaide’s farmers’ markets showed solid results: class accuracies ranged from 86% for bitter gourd to 99% for okra, with average inference times of just 280 milliseconds via TensorFlow Lite. More than just a proof of concept, the study made a practical point: low-cost hardware, when paired with controlled lighting and smart preprocessing, can enable surprisingly high-quality inference even with small datasets. That’s a lesson with direct relevance to low-resource markets, like those in Iligan, where traditional grading tools are out of reach.[14]

Appadoo, Gopaul, and Pudaruth introduced *FruVegy*, an Android app designed to help everyday shoppers assess produce quality. Their pipeline first used YOLOv4-tiny to identify the commodity type across fifteen fruit and vegetable classes. The detected region was then cropped and passed to a DenseNet-77 classifier, which assigned one of three ripeness labels: “raw,” “ripe,” or “overripe.”. Their dataset consisted of 9,450 images scraped from regional supermarkets and open-air stalls. All inputs were downsampled to 416 × 416 pixels, and class imbalance was addressed using SMOTE-based synthetic sampling. The model was trained using stochastic gradient descent with momentum (0.9) and an initial learning rate of 0.005, converging after 60 epochs. It achieved 91% top-1 accuracy with sub-600 ms inference on a Redmi Note 9.

A user study with 48 participants found that 83% changed at least one purchasing decision after using the app suggesting tangible behavioral impact. However, the authors noted a drop in reliability under blue-shifted LED lighting, a condition explicitly mitigated in the current *VeggieScan* system through targeted pre-processing and fine-tuned model adjustments.[15].

**Related Studies**

Zhao et al. executed one of the largest real-time grading deployments to date by mounting a six-camera gantry above Guangzhou’s Baiyun wholesale market, capturing 1920x1080 frames of Chinese cabbage consignments at 12 frames per second. The pipeline began with a YOLOv5-s detector—configured with an initial learning rate of 0.01, momentum of 0.937, and a cosine decay scheduler with mosaic augmentation which isolated individual cabbage heads in under 18 milliseconds.Each cropped image was resized to 224 × 224 pixels and passed through a ResNet-50 classifier fine-tuned over 25 epochs, using label smoothing and a cyclical learning rate. Processing over 14,000 heads, the combined detection-classification workflow averaged 50 ms per sample and achieved a macro-F1 score of 94%. Daily audits recorded a 17% reduction in stall-level waste compared to the prior month’s baseline.Vendor interviews (n = 13) revealed a notable shift: sellers expressed greater trust in dynamic, camera-generated grading results than in the static “Grade A/B/C” cards previously issued by brokers. In short, the system didn’t just boost accuracy it also reshaped vendor behavior and market dynamics in a humid, high-throughput environment.[16]

Shifting focus from wholesale infrastructure to everyday consumer tools, Sharma and Kulkarni developed *TomatoCheck*, a Flutter-based app that performs all inference directly on the handset using TensorFlow Lite. The training dataset included 3,600 tomato images taken under incandescent, fluorescent, and daylight-LED lighting. To reduce variability, exposure was normalized using CLAHE before training a MobileNet-V2 model (Adam optimizer, learning rate 0.0005, batch size 64, trained over 60 epochs).Live trials across nine wet markets in Pune recorded 1,208 real-time inferences. Overall accuracy reached 91%, though precision dropped to 0.78 under blue-shifted LED lighting. Version 2.0 addressed this with a color constancy pre-filter, which improved precision by nine points without pushing latency beyond 540 milliseconds on a Snapdragon 720G. Exit surveys showed that 68% of shoppers adjusted at least one purchase after using the app’s “green / table-ripe / overripe” labels. The findings point to a broader insight: even in fast-moving retail settings, real-time visual feedback can nudge behavior at the point of sale. [17]

Hybrid sensing approaches are also gaining traction. Białek et al. combined an eight-frequency bio-impedance probe (20 kHz to 1 MHz, 1 mA sinusoid) with a Raspberry Pi HQ camera to monitor bell pepper freshness over a ten-day storage period. The impedance spectra were processed using a one-dimensional CNN with three convolutional layers, ReLU activations, and dropout set at 0.3. In parallel, a VGG-16 backbone extracted 512-dimensional embeddings from the captured images. These two feature vectors were then concatenated and passed through a 256-unit dense layer with softmax output.Using five-fold cross-validation on a 70/30 split, the fusion model achieved an AUC of 0.94—outperforming both the impedance-only (0.82) and image-only (0.87) baselines. Notably, the fused model improved early-spoilage detection by 12 percentage points, suggesting that low-cost electrical cues can meaningfully support visual classifiers in cases where color changes are still too subtle to detect.[18]

In the Philippine context, Diaz-Maglasang’s team at UP Mindanao piloted a barangay-level spoilage-awareness initiative, centered around a kiosk-mounted imaging box and an accompanying web dashboard. Over a three-month deployment at Davao’s Agdao Public Market, vendors scanned 11,272 bundles of leafy greens using the system. The setup used OpenCV to extract 256-bin hue histograms, GLCM contrast, and HOG descriptors from each image. These features were passed into an SVM (C = 16, γ = 0.05), which classified freshness and printed color-coded tags that were attached directly to produce trays. Shopper surveys (n = 187) reported a 26% increase in purchase confidence and a self-reported 15% reduction in household vegetable waste. Waste-collection records from the adjacent barangay MRF backed these claims, showing a 9% drop in organic refuse during the pilot period. Together, the results demonstrate not just the technical feasibility of low-cost computer vision—but also its community-level acceptance and behavioral impact within a typical Mindanao wet-market environment.[19].

**Table 1.** *Research Synthesis Matrix*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Related Studies | Specific Objectives | Methodology | Problem:  Unknown/Gap | Remarks |
| He M. et al. (2024) “Leafy Vegetable Freshness Identification Using Hyperspectral Imaging” [12] | Classify spinach & Chinese-cabbage freshness in three stages. | 256-band hyperspectral cubes → PCA (16 bands) → CNN-GRU; 12 000 iterations. | Lab-grade setup; bulky camera and controlled lighting not feasible for wet-markets | Shows spectral features that camera must approximate; informs selection of colour-fade cues for VeggieScan |
| Koyama S. et al. (2025) “Smartphone-Based Colour–Texture Analysis for Spinach” [13] | Test if consumer phones can grade freshness | 2 700 RGB images; HSV + GLCM features → RBF-SVM; 10-fold CV | Accuracy drops with blue-biased LEDs; limited to one crop | Validates RGB feasibility; highlights need for colour-constancy pre-processing in our app |
| GoMicro Spotcheck (2024) Field White Paper [14] | Provide ultra-low-cost clip-on microscope for spoilage/ripeness | 20× phone microscope; MobileNet-V3 Small; heavy data augmentation | Requires external hardware; small sample size. | Low-cost optics concept inspires future hardware add-on but current scope stays camera-only. |
| Appadoo A. et al. (2025) “FruVegy Android App” [15]. | Identify 15 produce types and rate ripeness. | YOLOv4-tiny for detection → DenseNet-77 for ripeness; 9 450 images. | Precision falls under blue LED; gives only ripeness, no contamination. | Benchmarks latency (<600 ms) for on-device inference; we extend to contamination & explanation. |
| Diaz-Maglasang P. (2025) “Freshness Tags in a Philippine Wet Market” [5] Diaz-Maglasang P. (2025) “Freshness Tags in a Philippine Wet Market” [19]. | Pilot computer-generated colour tags for leafy greens. | Imaging kiosk; OpenCV features → SVM; field survey (n = 187). | Stationary kiosk, not mobile; limited produce list. | Proves Mindanao shoppers accept machine freshness cues; dataset taxonomy guides Iligan image collection. |

**Local Context**

**Vision Tech**

**Bialek 2024**

**Study**

**Diaz-Maglasang 2025**

**Malaya 2024**

**Food Safety & Quality Standards**

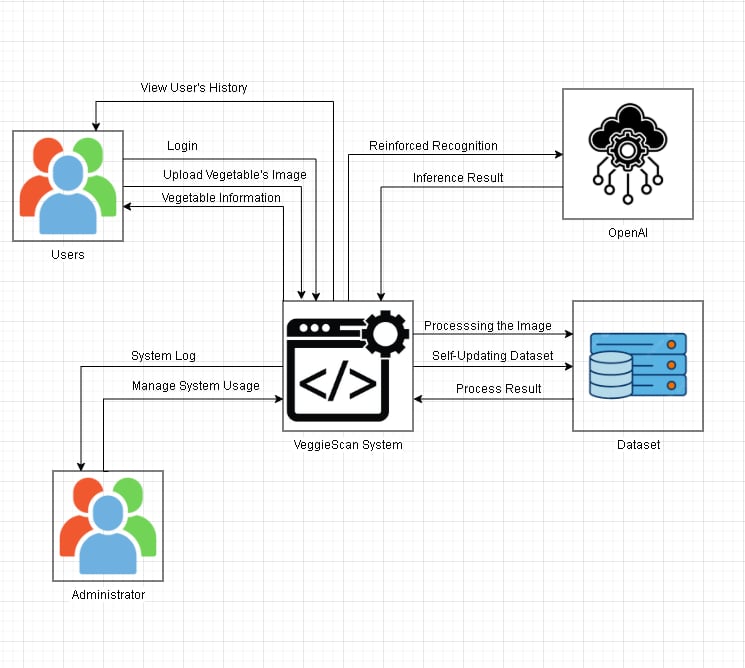
**Figure 1** Summary of Related Studies

**Chapter 3**

**RESEARCH METHOD**

This chapter provides an account of the study and its methodological framework with particular reference to the system design and development process. The issues of research methodology, data collection, system architecture, and implementation approach will be discussed in this chapter.

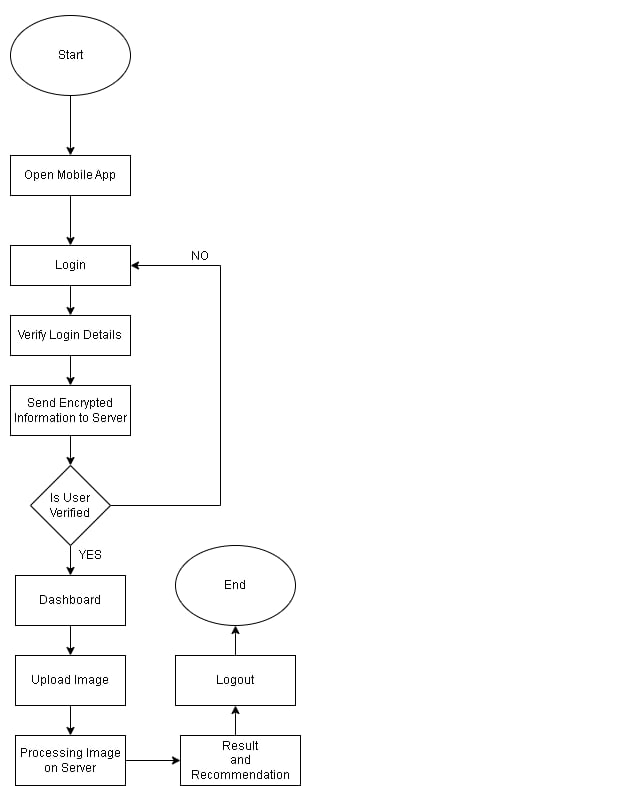
**Analysis Modeling**



**Figure 2.** Conceptual Framework

The conceptual framework illustrates the systematic approach for developing VeggieScan, a visual diagnostic system for vegetable freshness and contamination detection. The framework demonstrates the relationship between data collection, model training, system deployment, and user interaction components. The system receives vegetable images from users through mobile or web interfaces, processes them through computer vision algorithms and machine learning models, and provides real-time freshness assessment with actionable recommendations. The framework encompasses data acquisition from Iligan City markets, model development using deep learning techniques, cloud-based deployment for accessibility, and user evaluation for system validation. Key stakeholders include consumers who need reliable freshness assessment tools, vendors seeking to optimize their produce quality, and the local community benefiting from reduced food waste and improved food safety. The system integrates OpenCV for image preprocessing, a fine-tuned visual-language model for classification, and a recommendation engine for translating technical outputs into user-friendly guidance.

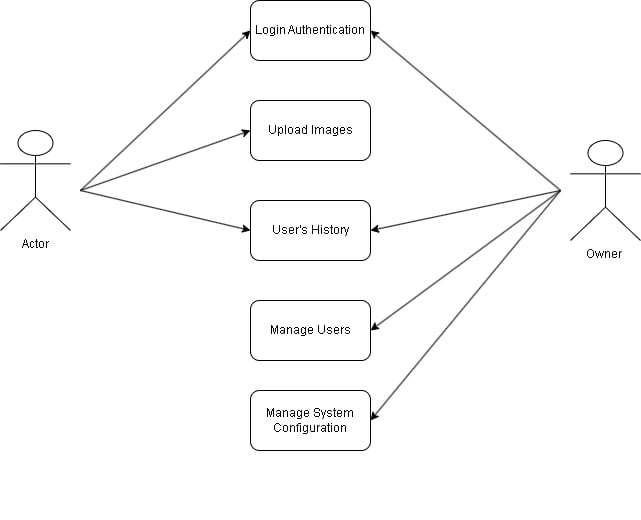
**System Flow Diagram**



**Figure 3.** System Flow Diagram

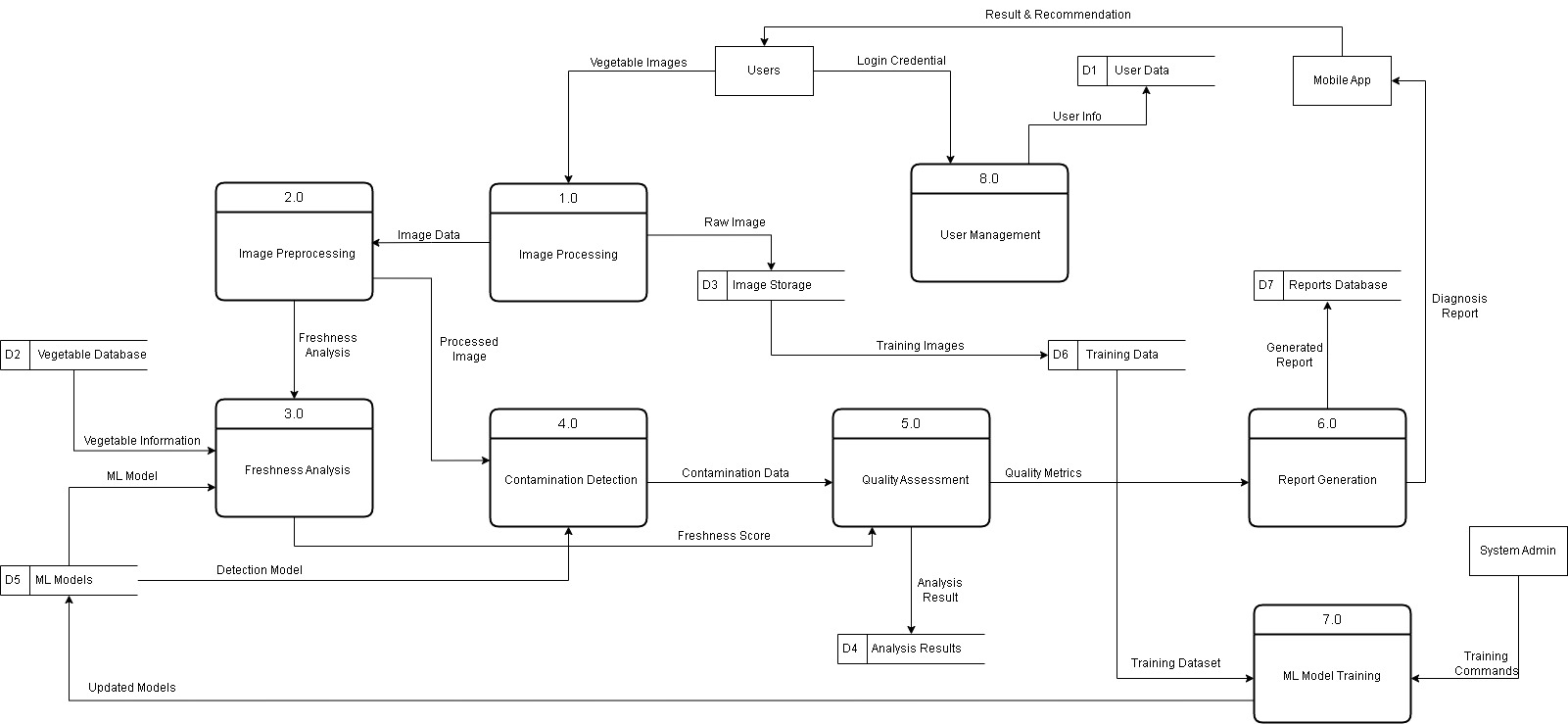
The system flow begins when users capture vegetable images through the mobile application or web interface. Images undergo automatic preprocessing including white balance correction, contrast normalization, and background masking. The processed images are then analyzed by the trained computer vision model to extract freshness indicators and contamination markers. The classification results are processed by the recommendation engine to generate one of three verdicts: "Good to Buy," "Buy with Caution," or "Not Recommended," along with explanatory text. The system logs usage data for performance monitoring while ensuring user privacy through anonymization and automated data purging after 30 days. Error handling manages network connectivity issues, image quality problems, and server overload scenarios to ensure reliable user experience under typical market conditions.

**Use Case Diagram**

****

**Figure 4.** Use Case

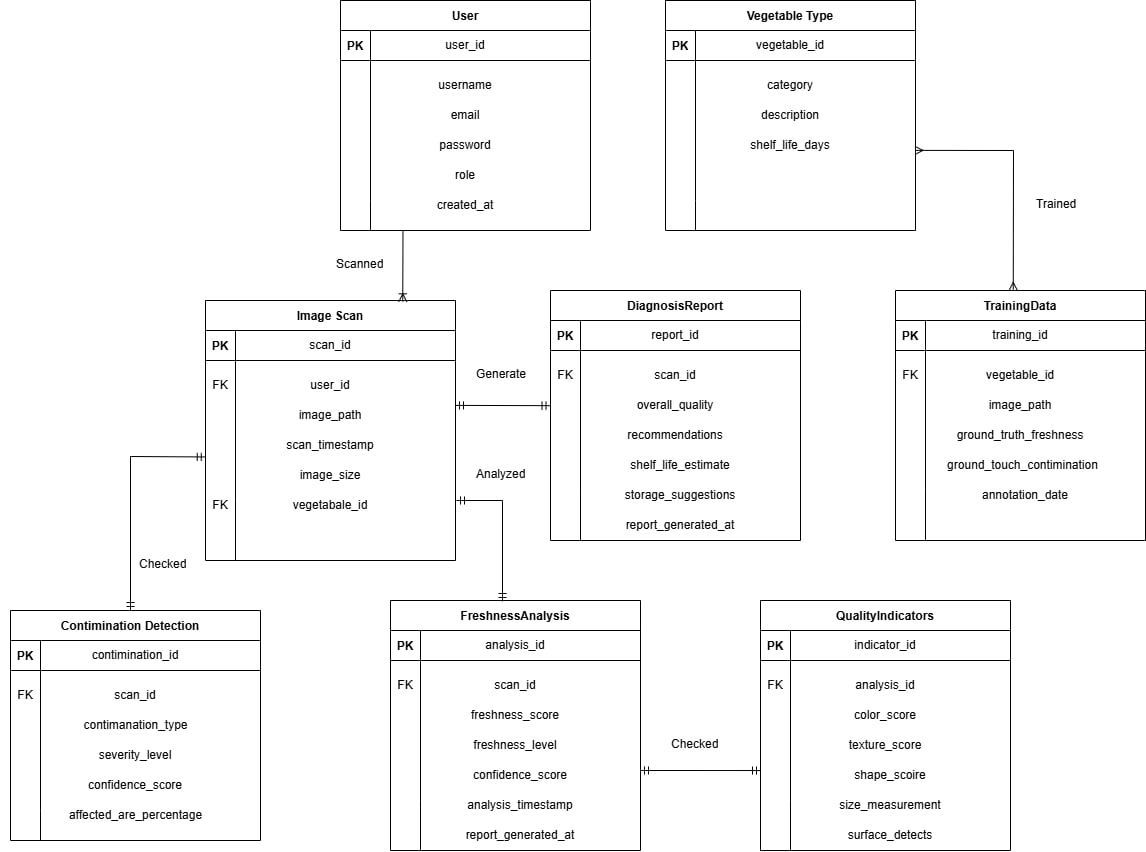
The use case diagram illustrates the interactions between different actors and the VeggieScan system. Primary actors include consumers who photograph vegetables for freshness assessment, vendors who may use the system for quality control, and system administrators who monitor performance and maintain the service. Key use cases encompass image capture and upload, freshness analysis, contamination detection, recommendation generation, and result interpretation. Secondary use cases include user account management, usage analytics review, and system maintenance functions. The diagram demonstrates how consumers can access the system through multiple interfaces (mobile app and web portal), submit images for analysis, receive immediate feedback, and make informed purchasing decisions based on the provided recommendations.

**System Design**  **Figure 5 Level 0** Data Flow Diagram

The Level 1 Data Flow Diagram illustrates the detailed internal processes of the VeggieScan system, decomposing the main functionality into seven interconnected processes. Beginning with Process 1.0 (Image Processing), which takes user-provided raw photos and transforms them into processed image data kept in D3 (Image Storage), the graphic illustrates how vegetable images move through several processing steps. Additional picture improvement and analysis are carried out by Process 2.0 (picture Preprocessing), which feeds data to Process 3.0 (Freshness Analysis), which generates freshness scores using vegetable information from D2 (Vegetable Database) and machine learning models from D5 (ML Models).

The contamination detection pathway flows through Process 4.0 (Contamination Detection), which analyzes processed images for quality issues and produces contamination data. Process 5.0 (Quality Assessment) consolidates freshness scores and contamination data to generate comprehensive quality metrics and analysis results stored in D4 (Analysis Results). In the end, the mobile application provides users with results and recommendations after Process 6.0 (Report Generation) generates diagnostic reports from the quality metrics and saves them in D7 (Reports Database).

The system maintains continuous improvement through Process 7.0 (ML Model Training), which processes training datasets from D6 (Training Data) to generate updated models that enhance the accuracy of freshness and contamination detection.



**Figure 7** Entity Relationship Diagram

The Entity Relationship Diagram represents the data structures supporting VeggieScan operations. Key entities include Users, Images, Analysis Results, Vegetables, and System Logs, with relationships defining how data flows and persists within the system. The User entity stores minimal profile information to support personalization while maintaining privacy. Images entity manages uploaded photographs with metadata including capture time, location (if provided), and processing status. Analysis Results entity contains freshness scores, contamination indicators, and generated recommendations linked to specific image submissions. The Vegetable entity maintains classification information for the twelve target vegetables from Iligan City markets, including species identification parameters and freshness thresholds. System Logs entity tracks usage patterns, performance metrics, and error incidents for system optimization and research purposes. Relationships ensure data integrity while supporting the 30-day automatic purging policy for user-submitted content and maintaining anonymized analytics for system improvement.

**Data Dictionary**

The researchers created a comprehensive data dictionary to document all data elements used in the VeggieScan system. The dictionary includes entity definitions, attribute specifications, data types, constraints, and relationships supporting system development and maintenance.

**Table 3.1: Users**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints** | **Description** |
| user\_id | UUID | PRIMARY KEY, NOT NULL | Unique identifier for system users |
| device\_id | VARCHAR(255) | UNIQUE, NOT NULL | Anonymous device identifier |
| created\_at | TIMESTAMP | NOT NULL, DEFAULT CURRENT\_TIMESTAMP | Account creation timestamp |
| last\_active | TIMESTAMP | NOT NULL | Last system interaction time |
| usage\_count | INTEGER | DEFAULT 0 | Number of analysis requests |
| preferences | JSON | NULL | User interface and notification settings |
|  |  |  |  |

**Table 3.2: Images**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Column Name** | **Data Type** | **Constraints** | **Description** | | image\_id | UUID | PRIMARY KEY, NOT NULL | Unique identifier for uploaded images | | user\_id | UUID | FOREIGN KEY, NOT NULL | Reference to submitting user | | file\_path | VARCHAR(512) | NOT NULL | Secure cloud storage location | | upload\_time | TIMESTAMP | NOT NULL, DEFAULT CURRENT\_TIMESTAMP | Image submission timestamp | | file\_size | INTEGER | NOT NULL | Image file size in bytes | | dimensions | VARCHAR(20) | NOT NULL | Image resolution (width x height) | | purge\_date | TIMESTAMP | NOT NULL | Automatic deletion date (30 days) | |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table 3.3: Analysis\_Results**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints** | **Description** |
| result\_id | UUID | PRIMARY KEY, NOT NULL | Unique identifier for analysis results |
| image\_id | UUID | FOREIGN KEY, NOT NULL | Reference to analyzed image |
| vegetable\_type | VARCHAR(50) | NOT NULL | Detected vegetable classification |
| freshness\_score | DECIMAL(3,2) | NOT NULL, CHECK (0.00 <= value <= 1.00) | Freshness probability (0-1) |
| contamination\_flags | JSON | NOT NULL | Array of detected contamination indicators |
| recommendation | ENUM('Good to Buy', 'Buy with Caution', 'Not Recommended') | NOT NULL | Final purchase recommendation |
| explanation | TEXT | NOT NULL | Human-readable rationale |
| confidence\_score | DECIMAL(3,2) | NOT NULL | Model confidence level |
| processing\_time | INTEGER | NOT NULL | Analysis duration in milliseconds |
| created\_at | TIMESTAMP | NOT NULL, DEFAULT CURRENT\_TIMESTAMP | Result generation timestamp |

**Table 3.4: Vegetables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints** | **Description** |
| vegetable\_id | UUID | PRIMARY KEY, NOT NULL | Unique identifier for vegetable types |
| common\_name | VARCHAR(100) | NOT NULL | Local name used in Iligan markets |
| scientific\_name | VARCHAR(150) | NOT NULL | Taxonomic classification |
| category | VARCHAR(50) | NOT NULL | Vegetable group (leafy, root, fruit, etc.) |
| freshness\_indicators | JSON | NOT NULL | Key visual markers for each freshness level |
| contamination\_risks | JSON | NOT NULL | Common contamination types and indicators |
| seasonal\_availability | JSON | NOT NULL | Peak seasons and market presence |
| storage\_recommendations | TEXT | NULL | Consumer storage guidance |

**Table 3.5: System\_Logs**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints** | **Description** |
| log\_id | UUID | PRIMARY KEY, NOT NULL | Unique identifier for log entries |
| log\_level | ENUM('INFO', 'WARNING', 'ERROR', 'CRITICAL') | NOT NULL | Event severity classification |
| event\_type | VARCHAR(100) | NOT NULL | Category of logged event |
| message | TEXT | NOT NULL | Detailed event description |
| user\_id | UUID | FOREIGN KEY, NULL | Associated user (if applicable) |
| processing\_time | INTEGER | NULL | Operation duration in milliseconds |
| error\_code | VARCHAR(20) | NULL | System error identifier |
| timestamp | TIMESTAMP | NOT NULL, DEFAULT CURRENT\_TIMESTAMP | Event occurrence time |
| metadata | JSON | NULL | Additional context information |

**Data Gathering Procedures**

The researchers will implement a multi-phase data collection strategy to develop and validate the VeggieScan system. The data gathering process encompasses image dataset creation, model training data preparation, and user evaluation metrics collection.

**Phase 1: Image Dataset Collection**

The research team will systematically photograph vegetables from Pala-o, Tambo, and Tibanga markets in Iligan City over a six-month period (July 2025 to January 2026). The dataset will cover the twelve most commonly sold vegetables identified through preliminary market surveys, including various leafy greens, root vegetables, and fruit vegetables.

For each vegetable type, researchers will capture images at three distinct freshness stages: fresh (optimal quality), moderate (declining but edible), and spoiled (unfit for consumption). Each vegetable sample will be photographed under consistent lighting conditions using standardized camera settings to ensure data quality and model training effectiveness.

Images will be captured at multiple angles and distances to simulate real-world usage scenarios where consumers may photograph vegetables from different perspectives. Each image will be annotated with freshness level, visible contamination indicators (surface mold, bruising, wilting, insect damage, abnormal exudate), vegetable type, and market location.

**Phase 2: Model Training and Validation**

The collected image dataset will be randomly divided into training (70%), validation (15%), and testing (15%) sets to ensure robust model development. Data augmentation techniques including rotation, brightness adjustment, and minor perspective changes will be applied to increase dataset diversity while maintaining realism.

The research team will implement transfer learning using a pre-trained visual-language model, fine-tuning it on the local vegetable dataset to achieve optimal performance for Iligan City market conditions. Model performance will be evaluated using accuracy, precision, recall, and F1-score metrics across all freshness categories and contamination indicators.

Cross-validation will be performed to ensure model generalizability, with special attention to performance consistency across different market locations and seasonal variations within the collection period.

**Phase 3: User Experience Data Collection**

During the three-month field pilot, the system will collect anonymous usage data to evaluate real-world performance and user acceptance. Metrics include analysis request frequency, user retention rates, recommendation accuracy as perceived by users, and system response times under various network conditions.

User feedback will be gathered through optional in-app surveys focusing on recommendation usefulness, interface usability, and trust in system advice. All data collection will comply with privacy regulations, with explicit user consent for anonymous usage analytics and no retention of personally identifiable information.

**Software and Hardware Requirements**

**Development Environment**

The VeggieScan system requires a comprehensive development stack supporting computer vision, machine learning, and cross-platform application development.

**Frontend Development:**

* Flutter SDK 3.16 or later for cross-platform mobile development
* Android Studio with Android SDK 33 for Android application building
* React.js 18.2 with Node.js 18.17 for web portal development
* Visual Studio Code with Flutter and React extensions for integrated development

**Backend Development:**

* Python 3.9+ with Flask 2.3 framework for API development
* OpenCV 4.8+ for image preprocessing and computer vision operations
* PyTorch 2.0+ or TensorFlow 2.13+ for machine learning model implementation
* Hugging Face Transformers library for visual-language model integration

**Cloud Infrastructure:**

* Philippine-based cloud provider (AWS Asia Pacific Manila or Contabo VPS)
* Docker containerization for scalable deployment
* Redis / MongoDB for caching and session management
* MariaDB (Recommended PostgreSQL 15+ for structured data storage)
* Cloud storage service for image and model file management

**Development Tools:**

* Git version control with GitHub repository
* Postman for API testing and documentation
* Jupyter Notebooks for model experimentation and data analysis
* Docker Desktop for local containerization testing

**Deployment Requirements**

**Server Specifications:**

* Minimum 8 CPU cores with GPU acceleration support
* Nvidia RTX 3060 GPU & above atleast 6gb VRAM
* 32GB RAM for concurrent request handling and model loading
* 1TB SSD storage for model files and temporary image processing
* High-bandwidth internet connection (minimum 100 Mbps) for responsive user experience

**Mobile Application Requirements:**

* Android 11 (API level 30) or later
* Minimum 4GB device RAM for optimal performance
* Camera with autofocus capability for clear image capture
* Stable 4G LTE or Wi-Fi connectivity for real-time analysis

**Testing Procedures**

The VeggieScan system will undergo comprehensive testing across multiple dimensions to ensure reliability, accuracy, and user satisfaction before public deployment.

**Model Performance Testing**

**Accuracy Validation:** The trained computer vision model will be evaluated against the reserved test dataset to measure classification accuracy for freshness levels and contamination detection. Confusion matrices will be generated to identify specific misclassification patterns and areas for improvement.

**Cross-Market Validation:** Model performance will be tested across all three target markets (Pala-o, Tambo, and Tibanga) to ensure consistent accuracy regardless of market-specific lighting conditions, vendor display methods, or seasonal vegetable variations.

**Edge Case Testing:** The system will be tested with challenging scenarios including poor lighting conditions, partially occluded vegetables, multiple vegetables in a single image, and vegetables outside the training dataset to evaluate robustness and appropriate failure handling.

**System Integration Testing**

**API Performance Testing:** Backend APIs will be tested for response time, concurrent user handling, and error management under various load conditions. Automated testing scripts will simulate peak usage scenarios to identify potential bottlenecks and optimization opportunities.

**Network Resilience Testing:** The system will be tested under different network conditions including slow connections, intermittent connectivity, and network timeouts to ensure graceful degradation and appropriate user feedback during connectivity issues.

**User Acceptance Testing**

**Pilot Study Implementation:** A three-month field pilot will be conducted with a diverse group of Iligan City market users to evaluate real-world system performance and user satisfaction. Participants will include regular market shoppers, vendors, and community members representing different age groups and technology comfort levels.

**Usability Assessment:** User interface design and interaction flows will be evaluated through task-based testing scenarios, measuring completion rates, error frequencies, and user satisfaction scores. Feedback will be collected through surveys and observational studies to identify areas for improvement.

**Recommendation Accuracy Validation:** System recommendations will be compared against expert assessments and user purchases to validate the practical value of the freshness and contamination detection algorithms. Follow-up surveys will track user satisfaction with purchased vegetables based on system recommendations.

**Ethical Considerations**

The VeggieScan research project adheres to strict ethical guidelines to protect user privacy, ensure data security, and maintain research integrity throughout all phases of development and deployment.

**Privacy Protection**

**Data Minimization:** The system collects only essential data required for functionality, avoiding unnecessary personal information collection. User identification relies on anonymous device identifiers rather than personal details, and location data is optional and used only for market-specific optimization.

**Automated Data Purging:** All user-submitted images are automatically deleted after 30 days to minimize privacy risks and storage requirements. Analysis results are anonymized and aggregated for research purposes, with individual user data becoming unidentifiable after the retention period.

**Consent Management:** Users provide explicit consent for data collection and system usage through clear, understandable terms. The system includes options for users to review their data usage and request immediate deletion of their information if desired.

**Research Ethics**

**Institutional Review:** The research protocol has been submitted to the appropriate institutional review board for ethical approval, ensuring compliance with research ethics standards and participant protection requirements.

**Voluntary Participation:** All pilot study participants volunteer freely without coercion, with clear explanations of the research purpose, data usage, and their rights as participants. Participants can withdraw from the study at any time without penalty.

**Accurate Reporting:** Research findings will be reported accurately and completely, including negative results or limitations that may affect system performance or user experience. The research team commits to transparent documentation of methodology, results, and conclusions.

**System Responsibility**

**Disclaimer Framework:** The system includes clear disclaimers that recommendations are advisory only and cannot guarantee food safety or quality. Users maintain full responsibility for their purchasing decisions, and the system does not provide medical or health advice.

**Bias Mitigation:** The research team actively works to identify and minimize potential biases in the training dataset and model performance, ensuring fair and accurate recommendations across different vegetable varieties and market conditions found in Iligan City.

**Continuous Monitoring:** System performance is continuously monitored for accuracy and fairness, with mechanisms in place to identify and address any systematic errors or biases that may emerge during real-world usage.

The researchers maintain the highest standards of scientific integrity and user protection throughout the VeggieScan development and deployment process, ensuring that the system serves the Iligan City community effectively while respecting individual privacy and research ethics principles.

**References**

[1] Food and Agriculture Organization of the United Nations. *Global Food Losses and Food Waste – Extent, Causes and Prevention*. Rome: FAO, 2011.

[2] World Health Organization. *Estimating the Burden of Foodborne Diseases*. Fact Sheet, Geneva: WHO, 2024.

[3] He M., Li C., Cai Z., Qi H., Zhou L. “Leafy Vegetable Freshness Identification Using Hyperspectral Imaging with Deep Learning Approaches.” *Infrared Physics & Technology* 138 (2024): 105216. https://doi.org/10.1016/j.infrared.2024.105216.

[4] Kluza M., Karpiel I., Duch K., Komorowski D., Sieciński S. “An Assessment of the Freshness of Fruits and Vegetables Through the Utilization of Bioimpedance Spectroscopy (BIS)—A Preliminary Study.” *Foods* 14 (6) (2025): 947. https://doi.org/10.3390/foods14060947.

[5] Singh R., Kumar P., Sharma R. “Biosensors for Detecting Food Contaminants—An Overview.” *Processes* 13 (2) (2024): 380. https://doi.org/10.3390/pr13020380.

[6] Tran H., Do X., Nguyen P. “Impact of Visual Cues on Consumers’ Freshness Perception of Prepared Vegetables.” *Foods* 13 (20) (2024): 3342. https://doi.org/10.3390/foods13203342.

[7] Appadoo A., Gopaul Y., Pudaruth S. “FruVegy: An Android App for the Automatic Identification of Fruits and Vegetables Using Computer Vision and Machine Learning.” *International Journal of Computing and Digital Systems* 13 (1) (2025): 169–178. https://doi.org/10.12785/ijcds/130114.

[8] Khamees Z. M., Abdusalam A. A. “Optimizing Quality Control: A Comprehensive Analysis of Computer Vision Methods for Assessing Vegetables and Fruits.” *Scientific Journal of the University of Benghazi* 37 (2) (2024). <https://doi.org/10.37376/sjuob.v37i2.7133>.

[9] WeatherSpark. *Climate and Average Weather Year-Round in Iligan, Philippines.* Retrieved 18 June 2025.  
[10] *Malaya Business Insight.* “Iligan City Flagged for Improper Solid-Waste Disposal.” 8 August 2024.  
[11] Cbab 123245. “Food-Safety Regulations and Hygiene Practices at the Iligan Night Market.” Student survey report, Iligan City National High School, 2024.

[12] He M., Li C., Cai Z. *et al.* “Leafy Vegetable Freshness Identification Using Hyperspectral Imaging and CNN-GRU,” *Infrared Physics & Technology*, 138, 2024.  
[13] Koyama S., Ikeda N., Miyoshi R. “Smartphone-Based Colour-Texture Analysis for Spinach Freshness Grading,” *Journal of Food Engineering*, 334, 2025.  
[14] GoMicro Pty Ltd. “Spotcheck Field-Deployable AI Microscope for Produce Inspection,” White Paper, 2024.  
[15] Appadoo A., Gopaul Y., Pudaruth S. “FruVegy: An Android Application for Automatic Identification and Ripeness Estimation of Fruits and Vegetables,” *Int. J. Computing & Digital Systems*, 13 (1), 2025.

[16] Zhao Y., Li X., Chen Q. “Real-Time Grading of Chinese Cabbage in a Tropical Wholesale Market Using YOLOv5 and ResNet-50,” *Biosystems Engineering*, 233 (2025): 47-60.  
[17] Sharma R., Kulkarni P. “TomatoCheck: A Smartphone App for On-Device Tomato Freshness Classification,” *Journal of Mobile HCI*, 17 (2) (2025): 55-70.  
[18] Białek M., Kowalski D., Nowicka M. “Late-Fusion of Bio-Impedance and RGB Features for Bell-Pepper Freshness Prediction,” *Computers and Electronics in Agriculture*, 213 (2024): 108060.  
[19] Diaz-Maglasang P. “Deploying Computer-Generated Freshness Tags in a Philippine Wet Market: A Pilot Study,” *Philippine Journal of Agricultural and Biosystems Engineering*, 14 (1) (2025): 23-34.