External Quality Inspection of Post-harvest Fruits using Computational Intelligence Methods

Luis Enrique Chuquimarca Jiménez

Facultad de Ingeniería en Electricidad y Computación - FIEC
Escuela Superior Politécnica del Litoral - ESPOL

This dissertation is submitted for the degree of
Doctor en Ingeniería Eléctrica, mención Control y Automatización

Guayaquil, January 2026

Director	**PhD. Boris X. Vintimilla** FIEC - ESPOL
Co-Director	**PhD. Sergio A. Velastin** School of EECS - Queen Mary University of London Dept. of Computer Science - University Carlos III
Thesis Committee	**PhD. Angel Sappa** FIEC - ESPOL **PhD. Miguel Realpe** FIEC - ESPOL **PhD. Jonathan Coronel** FIMCP - ESPOL **PhD. Xavier Soria** FIE - ESPOCH **PhD. Marco Mora** DI - UCM

This document was typeset by the author using LaTeX2e.

Declaración Expresa

Yo, **Luis Enrique Chuquimarca Jiménez**, acuerdo y reconozco que: La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor. El estudiante deberá procurar en cualquier caso de cesión de sus derechos patrimoniales incluir una cláusula en la cesión que proteja la vigencia de la licencia aquí concedida a la ESPOL.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, secreto empresarial, derechos patrimoniales de autor sobre software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al autor que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 21 de enero del 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
LUIS ENRIQUE
CHUQUIMARCA JIMENEZ

Luis Enrique Chuquimarca Jiménez

Evaluadores

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
ANGEL DOMINGO SAPPA
.

PhD. Ángel Sappa
FIEC – Escuela Superior Politécnica del Litoral

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
MIGUEL ANDRES REALPE ROBALINO

PhD. Miguel Realpe
FIEC – Escuela Superior Politécnica del Litoral

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
JONATHAN RICARDO CORONEL LEON

PhD. Jonathan Coronel
FIMCP – Escuela Superior Politécnica del Litoral

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
XAVIER SORIA POMA

PhD. Xavier Soria
FIE – Escuela Superior Politécnica del Chimborazo

PhD. Marco Mora
DI – Universidad Católica del Maule

I would like to dedicate this thesis with deep love and gratitude to my wife, Diana Herrera, and my son, Dylan Chuquimarca, who are my daily inspiration and the driving force behind my constant efforts. To my parents, Luis Chuquimarca and Alicia Jiménez, who instilled in me the values of perseverance and responsibility, and to my sister Alicia Chuquimarca, for her unconditional support. I also extend this dedication to my cousin Juan Chuquimarca and my nephews and niece Jana, Carlos, and Derek Chuquimarca, as well as to the rest of my family and my wife's family, for being a fundamental pillar in my life.

Finally, I dedicate this achievement to God, for granting me the strength, clarity, and perseverance needed to complete this important stage of my life.

Acknowledgements

And I am sincerely grateful to my thesis advisor, Boris Vintimilla, my co-advisor, Sergio Velastin, and the doctoral program coordinator, Ángel Sappa, for their invaluable academic and scientific guidance throughout this research journey. My appreciation also extends to the Dean of my Faculty at UPSE, Washington Torres, whose support was fundamental in this process.

I would also like to express my gratitude to my colleagues and collaborators at ESPOL and UPSE, whose encouragement, constructive feedback, and shared commitment to academic excellence greatly enriched this work. Their companionship and support during challenging moments strengthened my motivation and perseverance.

Finally, I acknowledge both institutions, ESPOL and UPSE, for providing the academic environment, resources, and opportunities that made this doctoral research possible. Their commitment to fostering research and innovation in higher education has been essential for the successful completion of this thesis.

Abstract

This dissertation addresses the challenge of automating post-harvest fruit quality inspection through the application of convolutional neural networks (CNNs) and deep learning models, with a focus on ripeness, defects, and deformities. A systematic review establishes the state of the art, identifying both advances and persistent gaps in CNN-based classification, including dataset limitations, lack of alignment with international grading standards, and insufficient robustness under real-world variability.

Building on this foundation, the study investigates banana ripeness classification under varying illumination, demonstrating that although models achieve high accuracy on pristine data, their generalization is significantly impaired under lighting changes. Gamma-based augmentation improves robustness, yet model performance remains architecture-dependent, underscoring structural limitations and the need for more resilient approaches.

Defect detection in apples and mangoes is examined through real and synthetic datasets, leveraging RGB, silhouette, and spectral-band imagery. The introduction of a Multi-Input architecture enhances discriminative power by integrating complementary representations, achieving superior accuracy compared to traditional Single-Input configurations. Furthermore, the exploration of spectral imaging highlights the potential of specific wavelengths for improved defect visibility.

The study also introduces a comprehensive dataset for fruit deformity classification, combining real, synthetic, and silhouette-based images of apples, mangoes, and strawberries. Comparative evaluations of standard and lightweight CNN models reveal the effectiveness of MobileNetV2, particularly in Multi-Input architectures, for capturing morphological variations.

Overall, the contributions of this research include the development of robust datasets, the proposal and evaluation of novel training strategies, and the validation of architectures capable of addressing key challenges in fruit quality inspection. The findings provide critical insights for advancing automated grading systems, highlighting trade-offs between robustness, efficiency, and scalability, while laying the groundwork for future exploration of transformer-based models, spectral imaging, and real-world deployment scenarios.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

B	denotes the base configuration trained (ViT)
FN	refers to false negatives
FP	refers to false positives
G	Green
I	original dataset images
R	Red
T	illumination map
TN	refers to true negatives
TP	refers to true positives
$I0$	is the unaltered dataset
$I1\ to\ I7$	incremental illumination variations defined by Γ
L	is the number of layers
$M0$	is the initially trained on datasets
N	is the number of images evaluated
P	is the total number of trainable parameters
t	is the time to process

Greek Symbols

Γ multiple illumination variants

γ produces brighter renderings

ρ a penalty factor

Superscripts

in input

out output

Subscripts

l is the layer

Other Symbols

$\hat{T}$ initial illumination map

$G^{(t+1)}$ auxiliary gradient variable

n_l^{in} is the input units of layer

n_l^{out} is the output units of layer

$T^{(t+1)}$ illumination estimate

$T_{\text{inference}}$ is the average inference time per image

$u^{(t+1)}$ penalty parameter

Acronyms / Abbreviations

AMS Agricultural Marketing Service

ANN Artificial Neural Network

BP Backpropagation

CNN Convolutional Neural Network

DCGAN Deep Convolutional Generative Adversarial Network

DL Deep Learning

EEC European Economic Commission

FAO Food and Agriculture Organization

FWHM Full Width at Half Maximum

GLCM Gray-Level Co-occurrence Matrix

GPU Graphics Processing Unit

HOG Histogram of Oriented Gradient

HSV Hue, Saturation, Value

LIME Low-light IMage Enhancement

LSTM Long Short-Term Memory

MHA Multi-Head Attention

GAN Generative Adversarial Networks

MLP Multi-Layer Perceptron

NIR Near-Infrared

OECD Organization for Economic Cooperation and Development

PSO Particle Swarm Optimization

SAM Segment Anything Model

SGD Stochastic Gradient Descent

SVM Support Vector Machine

TODN ThresholdedReLU Orthogonal Layer Weight Regularized DenseNet

UNECE United Nations Economic Commission for Europe

USDA United States Department of Agriculture

Vis Visible

ViT Vision transformer

WHO World Health Organization

Chapter 1

Introduction

1.1 Motivation

The fruit industry is increasingly relying on quality inspections using automated industrial equipment. External quality inspections of fruits occur in the post-harvest stage and focus on parameters such as ripeness, deformities, and defect detection. This situation demands more effective, accurate, and reliable processes. Globally, companies are exploring new technologies to improve these processes, such as the application of artificial vision techniques and computational intelligence methods. These technologies use image analysis to extract various parameters, such as ripeness, deformities, and defects, for grading fruits.

This trend is particularly significant in the context of food losses: fruits and vegetables suffer disproportionately high post-harvest waste, with global loss rates estimated between 28 % and 55 % of total production [91]. In many developing regions, such losses contribute directly to economic hardship and food insecurity; on a broader scale, they translate into a substantial waste of energy, water, labor, and financial resources. The Food and Agriculture Organization (FAO) similarly reports that up to one third of all food produced globally is lost before consumption [96]. These figures highlight the urgent need for automated, objective, and scalable inspection systems capable of improving post-harvest quality assessment and reducing losses along the supply chain.

In response to this need, several image datasets have been developed to support the training and evaluation of deep learning models for fruit quality inspection. Specifically, a dataset comprising four levels of banana ripeness, including both real and synthetic images, has been released to address ripeness classification tasks (https://github.com/luischuquim/BananaRipeness). In addition, a dataset of apple and mango images labeled as healthy or defective has been constructed to support defect detection (https://github.com/luischuquim/Healthy-Defective-Fruits). Similarly, a dataset containing images of strawberries, apples, and mangoes exhibiting structural deformations has been created to enable deformation analysis (https://github.com/luischuquim/Deformed-fruits). These datasets have been made publicly available to facilitate reproducibility and further research in automated fruit inspection.

The integration of these datasets with modern deep learning techniques enables the development of automated inspection systems targeting key computer vision tasks relevant to fruit quality assessment, namely classification, defect detection, and deformation analysis. By leveraging convolutional neural networks and related architectures, such systems can accurately evaluate ripeness levels, identify surface defects, and detect shape irregularities under diverse imaging conditions. For example, banana ripeness can be classified using the dedicated ripeness dataset, while surface defects in apples and mangoes can be detected using the corresponding defect dataset. Likewise, the deformation dataset supports the identification of structural anomalies in strawberries, apples, and mangoes. Together, these task-oriented datasets provide a comprehensive foundation for improving the precision, robustness, and scalability of automated fruit quality inspection systems.

By inspecting fruit with computer vision and DL models, defective units are automatically identified and removed before they advance down the line, preventing cross-contamination of batches and significantly reducing waste and associated losses. These innovations not only promise to reduce waste and improve the overall quality of fruit products but also align with global sustainability goals by minimizing resource usage and maximizing yield. As the fruit industry continues to adopt these technologies, it stands on the brink of a significant transformation, where traditional methods give way to a more intelligent, data-driven approach to quality control.

Faced with the growing shortage of skilled labor in the fruit sector, coupled with the need to meet increasingly stringent quality standards and address sustainability and climate-change challenges, our project leverages the synergy between state-of-the-art DL models and high-fidelity datasets to deliver a comprehensive automated inspection solution.

In this context, the initiative elevates the precision and efficiency of high-throughput fruit processing by integrating deep-learning algorithms with cutting-edge hardware. This approach enables continuous, automated classification on the production line, ensuring consistent inspection and freeing human resources for higher-value tasks. In doing so, it improves the competitiveness and sustainability of the fruit industry, fostering smarter, more resilient supply chains aligned with contemporary environmental goals.

1.2 Research Objectives

1.2.1 General Objective

- Inspecting the external quality of fruits using computational intelligence methods evaluated with high-quality datasets for fruit classification.

1.2.2 Specific Objectives

- Create high-quality datasets comprising images of fruits to assess their external characteristics, taking into account parameters such as ripeness, deformities, and defects.
- Evaluate intelligent models for the precise classification of fruit ripeness, using relevant datasets and specific metrics to enable a detailed assessment of their maturation status.
- Evaluate intelligent models for the classification of fruit deformations, using relevant datasets and assessing their performance with appropriate metrics.
- Evaluate intelligent models for the classification of surface defects in fruits, using relevant datasets and appropriate metrics to ensure their performance.

1.3 Research Questions

Based on the objectives outlined, this study seeks to address the following research questions:

1. How can comprehensive, high-quality image datasets be constructed to accurately represent external fruit quality parameters such as ripeness, deformation, and defects?
2. To what extent do state-of-the-art deep learning models achieve precise classification of fruit ripeness when evaluated on these datasets using metrics such as accuracy, precision, recall, and F1 score?
3. How effectively can convolutional neural networks distinguish and classify morphological deformations in fruits, and which architectures yield the best performance under standardized evaluation protocols?
4. How robust are CNN-based approaches for detecting and categorizing surface defects in fruits, and what levels of sensitivity and specificity can be attained?

1.4 Research Methodology

This section presents the research methodology for this doctoral thesis; see Figure 1.1. Section 1.4.1 describes the acquisition and generation of real and synthetic datasets in multiple spectra for training, and testing computational models. Section 1.4.2 details the development of computational methods and architectures for classifying external fruit quality parameters such as ripeness deformation and defects. Section 1.4.3 reviews the metrics used to assess model performance, including accuracy, precision, recall, and F1 score.

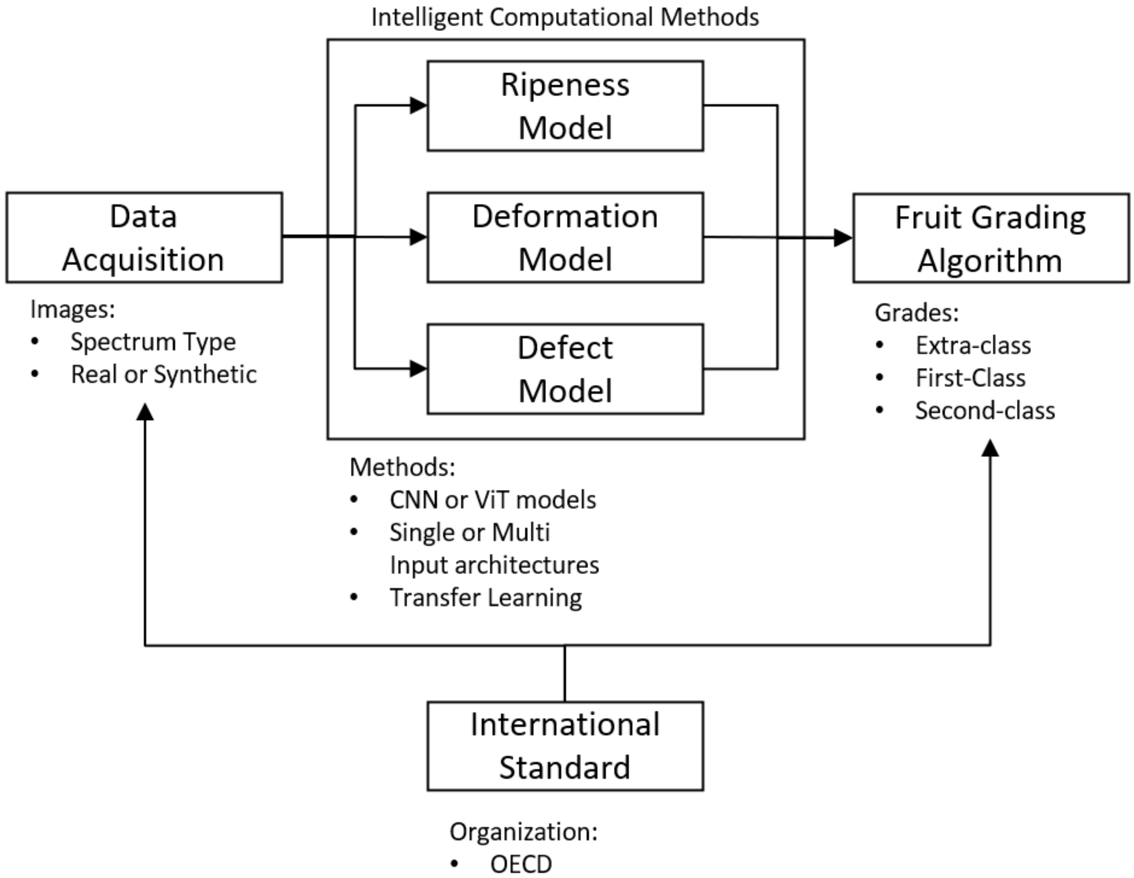


Fig. 1.1 General methodology proposed for this research.

1.4.1 Data Generation

Datasets focus on a single fruit and represent the quality parameters of ripeness, deformation, and defects. The fruits considered for this study are bananas, apples, mangoes, and strawberries, selected due to their national production relevance and the strong research background reported in the state of the art. Since each fruit presents various maturity stages, independent datasets are generated for each ripeness level. Likewise, distinct datasets are developed for each deformation grade. For the defect parameter, datasets include both healthy fruits and fruits showing different defect types; when specific defect categories must be identified, an additional dedicated dataset is produced for each category.

The fruits and corresponding evaluated parameters are selected based on physiological behavior and observable quality traits, whereby bananas are used for ripeness analysis, apples and mangoes for defect characterization, and apples, mangoes and strawberries for deformation assessment, as summarized in Table 1.1.

Table 1.1 Selection of fruits, associated evaluated parameters, and justification.

Fruit	**Parameter**	**Justification**
Banana	Ripeness	Climacteric fruit with well-defined biochemical and visual transitions; extensive state-of-the-art research enables reliable benchmarking for maturity classification.
Apple	Defects	Multiple visibly distinguishable defects (bruising, fungal spots, scabbing); well documented in literature, suitable for multi-class defect modeling.
Mango		Broad diversity of external defects (anthracnose, pitting, sap burn); strongly supported by previous studies, suitable for defect classification.
Apple	Deformations	Geometry allows clear shape deviations; morphological traits support categorical grading of deformation severity.
Mango		Structural form presents distinct geometric variations; supports reliable deformation categorization.
Strawberry		Natural irregular morphology facilitates geometric-based grouping of deformation levels.

Model evaluation relies on real datasets captured with cameras operating in the visible and near infrared, as well as refined images from public repositories. Synthetic datasets are produced using tools such as Unreal Engine, DALL E Mini and Stable Diffusion. To enhance dataset quality, image preprocessing techniques are applied, including noise reduction and contrast enhancement to facilitate parameter extraction. Segmentation isolates fruits from the background so that analysis focuses exclusively on ripeness, deformation and defects. Finally, data augmentation techniques such as rotation, flipping and brightness adjustment expand dataset diversity, strengthen model robustness and reduce the risk of overfitting.

1.4.2 Development of Intelligent Models

Computational intelligence methods are applied to extract key external quality features such as ripeness, deformation and defects for accurate fruit classification. Model selection is guided by a comprehensive review of state of the art architectures, ensuring each network is well suited to the specific grading task.

To optimize performance, this study employs transfer learning with single input and multi input designs. Models are first trained on synthetically generated images to establish robust feature representations and are then fine tuned on real datasets acquired in the visible and narrowband near infrared spectra. This strategy leverages pretrained weights to improve generalization and increase accuracy in classifying external quality parameters (see Figure 1.2).

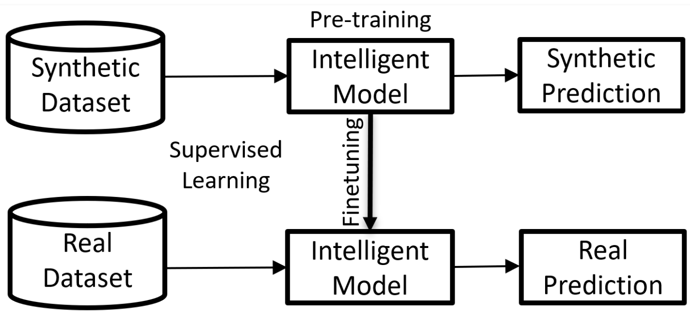


Fig. 1.2 Transfer Learning process.

This thesis employs a single-input convolutional neural network architecture to classify external quality parameters of fruits, harnessing its computational efficiency and rapid processing capabilities. By processing a single image modality, this approach simplifies model evaluation and hyperparameter tuning (see Figure 1.3).

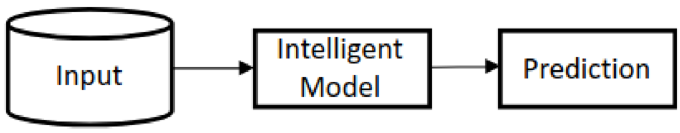


Fig. 1.3 Single-Input architecture.

Another methodology component is the use of multi input architectures that integrate diverse data sources into a unified neural network. Each input channel processes images from a distinct spectral domain, such as visible and near infrared, enabling complementary feature extraction and improving classification accuracy of external quality parameters in fruits. This configuration also supports the inclusion of dedicated preprocessing outputs in separate branches, for example segmentation masks or binarized images, which provide additional spatial and textural cues. By combining multiple inputs, the model leverages varied information and tailored preprocessing to enhance its predictive performance in fruit quality assessment (see Figure 1.4).

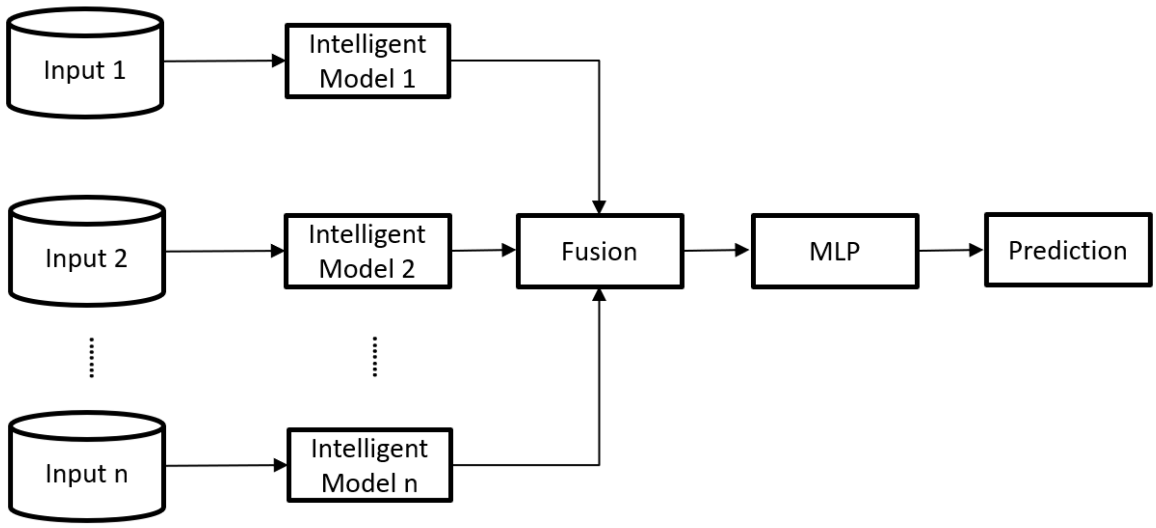


Fig. 1.4 Multi-Input architecture.

1.4.3 Evaluation Metrics

The essential metrics for evaluating intelligent models working with image datasets are chosen following a comprehensive review of the state-of-the-art fruit quality inspection cases (see Table 2.9). Consequently, the most commonly used metric is accuracy, provided the dataset is as balanced as possible. This is followed by precision, which measures the proportion of correct predictions relative to the total predictions; recall, which assesses the model's ability to detect true positives, i.e., the defect detection rate; and the F1 score, which combines precision and sensitivity to provide a balanced measure of model performance.

1.5 Required Resources

The DL models employed in this study are implemented in Python using established deep learning frameworks, including TensorFlow, PyTorch, and the Keras library. To reduce training time and ensure efficient computation, the experiments are conducted on a high-performance workstation equipped with an Intel I9-9820X processor, 128 GB of RAM, and a graphics processing unit (GPUs), specifically the NVIDIA RTX 2080 Ti. Furthermore, dedicated camera equipment is utilized to capture high-quality real images, supporting the construction of datasets required for the training and validation of the models.

Critical resources for this study include the generation of synthetic images through advanced text-to-image models such as DALL·E, as well as image-to-image approaches based on generative adversarial networks (GANs). The methodology also incorporates image segmentation techniques to refine structural details and employs diffusion-based models, including Stable Diffusion, to enhance stability and improve visual quality. These tools collectively strengthen the experimental framework, significantly contributing to the robustness and success of the study.

1.6 Contributions and Outline

The outline and contributions of this dissertation are:

- **Chapter 2.** Provides a comprehensive review of the application of CNN models in post-harvest fruit classification, aligned with international standards of external quality inspection. This chapter constitutes, to the best of our knowledge, the first systematic review that analyzes CNN-based fruit grading approaches in detail, covering studies from 2015 to the present. It summarizes the main datasets, parameters, and evaluation metrics employed, and identifies research gaps, particularly regarding dataset size, generalization, and robustness against real-world variability.

- **Chapter 3.** Investigates the robustness of CNN and transformer-based models for banana ripeness classification under variable illumination conditions. It introduces lighting variations into the dataset to test model stability and robustness, and performs a comparative analysis of architectures considering accuracy, number of parameters, inference time, and computational efficiency, offering criteria for selecting the most effective models for real-world deployment.

- **Chapter 4.** Focuses on defect detection in apples and mangoes using real and synthetic datasets. It introduces improvements over previous studies by exclusively employing real images and incorporating silhouette pre-processing to enhance defect visibility. A Multi-Input architecture is proposed to integrate RGB and silhouette images, achieving superior results compared to Single-Input approaches. Additionally, an extended study evaluates apple defect classification under different spectral ranges (visible and 660 nm), producing a publicly available dataset and analyzing the trade-offs of Single-Input and Multi-Input schemes for defect classification.

- **Chapter 5.** Addresses the classification of fruit deformities through the development of three distinct datasets: real, synthetic, and silhouette-based. This chapter evaluates lightweight and standard CNN architectures (MobileNetV2, VGG16, and CIDIS), demonstrating the advantages of lightweight networks for deformity detection. It further examines Single-Input and Multi-Input training approaches, highlighting the benefits of complementary representations (RGB and silhouette) for more precise shape-based classification. The datasets generated are made publicly available, contributing valuable resources to the research community.

- **Chapter 6.** Presents the publications derived from this dissertation, summarizing their content, impact, and the dissemination of the main contributions through international conferences and a scientific journal.

The contributions mentioned above have been presented at international conferences and published in a scientific journal. More details about these publications can be found in Chapter 6.

The thesis is organized as follows. Chapter 2 summarizes relevant works related to fruit quality inspection using CNN models. Chapter 3 describes the methodology and results of assessing robustness for banana ripeness classification under varying illumination. Chapter 4 addresses defect detection in apples and mangoes, including spectral analysis. Chapter 5 focuses on deformity classification using multi-source datasets and input architectures. Finally, Chapter 6 compiles the publications derived from this research.

Chapter 2

Literature Review

2.1 Introduction

This section is grounded in a comprehensive journal review titled "A review of external quality inspection for fruit grading using CNN models," which serves as the foundation for the analysis presented here. The review critically examines state-of-the-art convolutional neural network applications for post-harvest external assessment of fruit quality, focusing on key parameters such as ripeness, deformation and defects. Emphasis is placed on alignment with international grading standards, dataset characteristics and comparative model performance, providing a structured framework for the detailed discussion that follows.

Nowadays, quality control is critical for companies worldwide, ensuring that products meet consumer expectations throughout every stage of production [58]. Food producers adhere to international quality standards to maintain product compliance and market competitiveness [142, 186]. Agricultural products such as fruits and vegetables play a vital role in providing balanced and healthy nutrition; thus, their quality must be consistently evaluated [4, 63, 118, 208]. Effective quality assessment directly influences the categorization of these products, with higher categories translating into greater economic value [97, 123]. Accurate categorization benefits producers by facilitating market entry and optimizing pricing strategies [185]. Consequently, reliable product inspections are essential globally, assisting in categorization, determining shelf life, and guiding optimal distribution practices [157, 194].

In agricultural production, each stage from cultivation to post-harvest significantly impacts global food security and sustainability [146, 204]. The cultivation stage establishes the basis for crop quality and yield, whereas the harvest involves collecting the products for subsequent processes [82, 159, 179]. The post-harvest stage, which includes quality inspections, storage, and marketing, presents substantial challenges for agricultural management. Effective post-harvest practices such as sorting products by size, shape, color, and quality criteria, assessing physical damages, controlling pests and diseases, proper refrigeration, packaging verification, and appropriate transportation are essential to preserve product integrity and extend shelf life. Inefficient management during this stage results in reduced food quality, diminished nutritional value, and decreased market worth, contributing significantly to global losses estimated at approximately 1.3 billion tons annually, as reported by the FAO [102].

To successfully market their products, fruit industries must meet external quality standards related to color, shape, size, and defects, as these are crucial indicators during post-harvest selection [210]. Efficient, precise, and automated processes are essential for continuously classifying harvested fruits, emphasizing their external quality and reducing the need for invasive inspections. Traditional manual visual inspections often lead to inconsistent, subjective, time-consuming, and inaccurate evaluations [19]. Addressing these limitations can significantly enhance profitability for producers while decreasing production time. Consequently, implementing advanced technologies, such as artificial vision techniques, emerges as an effective solution for improving quality assessment in the agricultural sector.

Artificial vision encompasses image-based techniques to identify key parameters such as shape and color, particularly useful in agricultural applications during harvest and post-harvest stages [154]. These systems automatically extract relevant features from images and learn underlying patterns through machine learning [181]. DL, a subset of machine learning characterized by deep neural networks, achieves notably higher accuracy compared to traditional approaches. Specifically, Convolutional Neural Networks (CNNs), which apply convolutional techniques to analyze images, have proven highly effective for visual assessments [56, 61, 106, 117, 191, 222]. Consequently, employing CNNs for external fruit quality inspection aligns with international standards and commercial regulations, offering a precise and reliable solution.

International standards for fruit categorization, published by trade regulatory organizations in the United States and Europe, establish unified criteria for external quality inspections [17]. These standards focus primarily on parameters such as ripeness, deformations, and defects, enabling fruit classification into distinct quality grades [83]. For instance, according to the Organization for Economic Cooperation and Development (OECD), first-class fruits meet high-quality criteria and command premium prices on the international market [36]. In comparison, second-class fruits represent medium quality, while third-class fruits are considered low quality and thus have a lower market value [40].

Automating fruit inspection processes requires alignment with international quality standards. Recent advancements in machine vision facilitate precise feature extraction from fruit images. DL techniques have demonstrated high performance and significant potential for accurately assessing fruit quality parameters to determine their corresponding grade or category.

2.2 Fruit Variety and Quality Standards

Fresh fruits are essential sources of carbohydrates, acids, vitamins, and minerals, contributing significantly to human health [105]. Additionally, fruits serve as raw materials for agro-industrial applications, including the production of oils, pharmaceuticals, dyes, and other derivative products [155]. Given their economic importance, companies continually seek to optimize fruit production and enhance post-harvest product quality. This section outlines the key characteristics distinguishing fruit varieties, emphasizing aspects critical to post-harvest evaluation.

During fruit development and maturation, species are broadly classified as climacteric or non-climacteric according to their ethylene production and respiration patterns [89, 100]. Climacteric fruits continue to ripen after harvest, exhibiting a pronounced rise in respiration and ethylene synthesis; examples include bananas, apples, mangoes, pears, avocados, peaches, plums, kiwis, papayas, and passion fruits [31, 144, 156]. In contrast, non-climacteric fruits, such as grapes, lemons, oranges, strawberries, raspberries, watermelons, pineapples, and other citrus species, show minimal changes in respiration and maintain low ethylene levels [14, 215].

This study examines both climacteric and non-climacteric fruits to account for the pronounced differences in their ripening behaviour and the resulting effects on external attributes such as deformation and visible defects. Climacteric species, such as bananas, apples and mangoes, continue to ripen after harvest, causing their surface characteristics to evolve and potentially reveal defects that were not apparent at picking. In contrast, non-climacteric fruits like strawberries reach full maturity on the plant and retain a stable external appearance once harvested.

Fruit quality can be assessed through invasive internal methods, which measure attributes such as sugar concentration and acidity by physically penetrating the fruit, or through non-invasive external methods that evaluate parameters like ripeness, deformation, and surface defects without compromising structural integrity. This work focuses on external inspection, pairing image-based observation with CNN models to preserve product quality while achieving reliable classification.

External quality attributes such as peel color and surface texture are closely related to these physiological behaviors [21, 158], and play a critical role in consumer acceptance [49].

Fruit quality assessments can be grouped into two main categories [144, 223] (see Figure 2.1):

- Internal quality encompasses flavor attributes (sweetness, acidity, bitterness, saltiness, astringency, aromatic compounds), textural properties (firmness, crispness, juiciness, tenderness, chewiness, fibrosity), nutritional composition (fats, carbohydrates, proteins, vitamins, minerals) and latent defects (internal cavities, water core, frost damage, rot). Because these characteristics generally require destructive or invasive testing, evaluations are limited to small sample sets.

- External quality involves measurable traits such as size (weight, volume, linear dimensions), shape (diameter–depth ratios), color (hue, uniformity, intensity), and visible defects (surface fungi, bruises, punctures, blemishes). These attributes lend themselves to non-invasive imaging techniques, which constitute the focus of this review, particularly when combined with CNN-based analytical models.

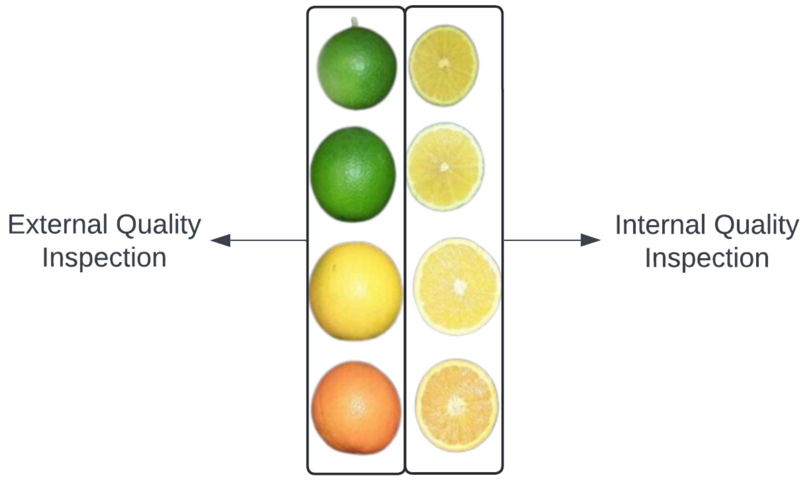


Fig. 2.1 External and internal quality inspection of fruits

This study focuses on external fruit quality assessment through noninvasive imaging methods that quantify ripeness, deformation, and surface defects without compromising structural integrity. Although invasive internal tests provide precise information on attributes such as sugar content and acidity, they damage the product and are impractical for large-scale inspection. Image-based external inspection offers a rapid and scalable alternative, capturing visual cues directly associated with freshness, appearance, and consumer acceptance. Continuous monitoring of these parameters safeguards food safety and quality while supporting the fruit industry's commitment to stringent international standards.

The United States Department of Agriculture (USDA) develops, revises, and enforces grade standards that govern the marketing of fresh fruits [198]. Although these standards apply to only certain commodities, they rely chiefly on visual assessment and sort produce into hierarchical categories U.S. Fancy, U.S. No. 1, U.S. No. 2, U.S. No. 3, U.S. Extra No. 1, and U.S. Extra Fancy that signal quality to buyers and consumers. Within packing facilities, inspectors from the Agricultural Marketing Service (AMS) perform routine checks along the packing line, sampling lots to confirm compliance with the stated grade. Their findings are relayed to management for prompt corrective action. To improve consistency, AMS equips field inspectors with digital cameras and networked computers, enabling rapid transmission of fruit images to quality experts at USDA headquarters for real-time consultation [86, 170].

The European Economic Commission (EEC) introduced international quality and marketing standards for fresh fruit in 1954. These guidelines were formalized in 1961 as the European International Standards, providing criteria for inspecting and certifying both imported and exported fruit. The system defines three quality categories: Extra Class, denoting superior quality; Class I, representing good quality; and Class II, indicating marketable quality [86].

The United Nations Economic Commission for Europe (UNECE) establishes an international grading scheme that assigns fruits to three quality tiers [55]. Extra Class denotes the highest grade, reserved for fruit displaying optimal appearance and only negligible superficial defects. First Class corresponds to good-quality produce with slight imperfections, uniform presentation, and the minimum level of ripeness required for marketability. Second Class includes fruit of acceptable quality that may exhibit visible defects yet still retains all essential characteristics for consumer use.

The Codex Alimentarius Commission, jointly established by the FAO and the World Health Organization (WHO), issues international standards that categorize fresh fruit into three quality tiers [107]. Extra Class represents the premium grade, encompassing fruit of superior quality that is free from notable defects and exhibits optimal ripeness, uniform shape, and pristine appearance. Class I, often referred to as First Class, includes good quality fruit that maintains the essential characteristics of the variety but may show slight, superficial imperfections such as minor skin blemishes or faint discoloration that do not compromise overall presentation or shelf life. Class II, or Second Class, permits more pronounced defects, including visible surface marks, shape irregularities, or slight bruising, provided the fruit remains structurally sound, safe, and fit for consumption. This grading framework facilitates transparent trade and ensures consistent quality benchmarks across international markets.

OECD also applies a three-tier grading system for fresh produce. Extra class represents the highest quality and imposes rigorous criteria on ripeness, morphology, and absence of

Table 2.1 Parameters used by fruit grading

Parameters/Fruits	**Apple**	**Orange**	**Banana**	**Lemon**	**Mango**	**Strawberry**
Ripeness	✓	✓	✓	✓	✓	✓
Deformations	✓	×	✓	×	✓	✓
Defects	✓	✓	✓	✓	✓	✓

defects: fruit must reach optimal maturity for the species, for example, apples should be fully ripe yet unwrapped, exhibit a uniform and attractive shape without deformities, and remain completely free of visible imperfections such as bruises, scars, discoloration, disease or insect damage [78]. First Class comprises good-quality fruit that meets strict visual and structural requirements but may display minor, superficial defects that do not affect overall appearance or storage potential. Second Class includes fruit with more pronounced defects, such as slight bruising or irregular shape, yet still retains the essential characteristics and edibility required for commercial distribution. This hierarchical framework ensures consistent quality benchmarks across international markets, facilitating transparent trade and safeguarding consumer expectations.

International grading standards emphasize ripeness, deformation, and defects, complemented by detailed guidelines for post-harvest handling, transportation, and storage to preserve quality and safety across the supply chain. This thesis adopts the OECD framework and concentrates on those three parameters as primary indicators for external fruit quality assessment. A comprehensive literature review identifies the essential grading criteria for each fruit species in alignment with these international standards. Table 2.1 consolidates the parameters ripeness, deformities, and defects mapped to the principal fruits examined. Notably, certain fruits do not require assessment of every parameter, particularly deformation, underscoring the need for species-specific inspection protocols.

This doctoral research targets automated fruit grading using key external parameters ripeness, deformation, and defects and focuses on four globally significant export commodities: banana, apple, mango, and strawberry. The study concentrates on meticulous post-harvest inspection of external quality (Figure 2.2), the stage at which fruits receive their final grading before entering commercial distribution. A survey of the current literature indicates that post-harvest quality assessment is often under-represented, with many investigations overlooking the specific challenges of external inspection required for market compliance. By addressing this gap, the present work seeks to develop targeted methodologies that enhance the reliability and efficiency of post-harvest grading, thereby supporting international trade standards and reducing quality-related losses.

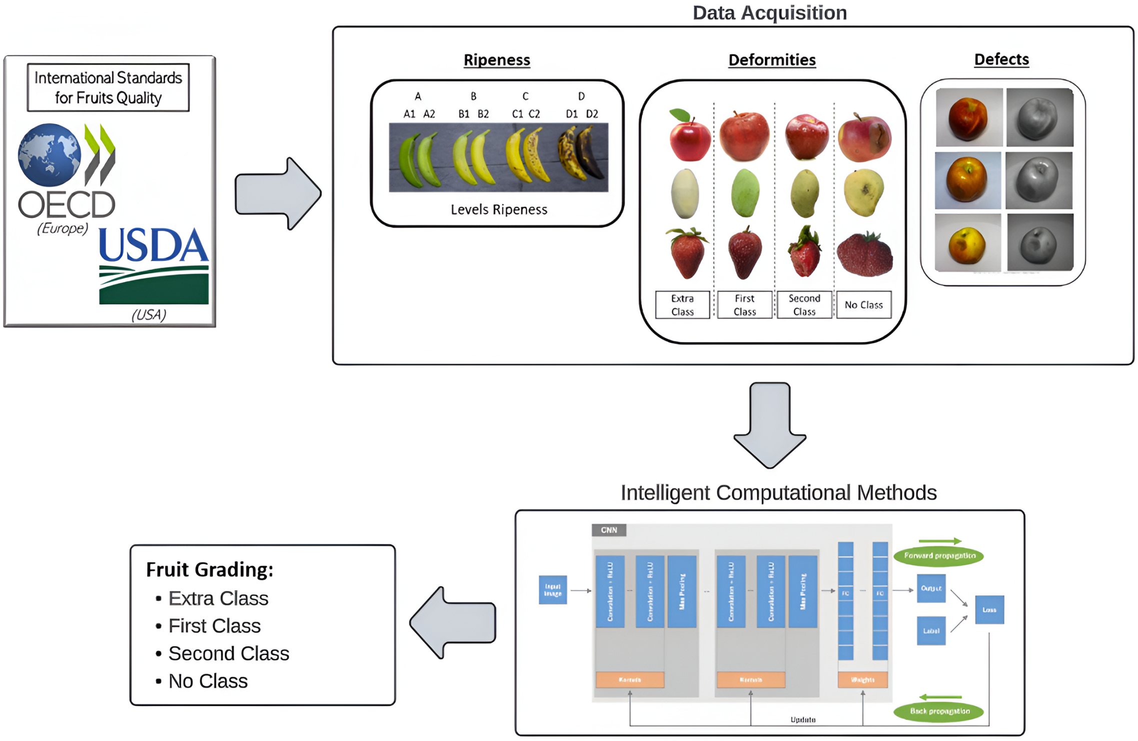


Fig. 2.2 Proposed external fruit quality inspection process.

In post-harvest inspection, ripeness, deformation, and surface defects constitute the critical external attributes that govern fruit quality assessment. Ripeness defines the optimal developmental stage and directly influences flavor, texture, and shelf life, thereby shaping consumer acceptance. Deformations reflect irregular growth or handling and detract from visual appeal, potentially undermining marketability. Surface defects such as stains, cracks, or bruises signal mechanical damage or disease that can compromise appearance, safety, and nutritional value. Computational intelligence methods that quantify these features enable precise grading into the OECD tiers (Extra Class, First Class, Second Class) and promote consistent quality in commercial distribution. The literature further underscores the importance of high quality datasets for training and validating CNNs, and highlights a research gap in the application of Transformer architectures to fruit classification a promising direction for achieving greater model robustness. Rigorous evaluation with appropriate performance metrics is essential to substantiate the effectiveness of any DL approach. Accordingly, this study offers a comprehensive review of state-of-the-art computational techniques for external fruit quality inspection, with particular attention to dataset robustness, model selection, and evaluation protocols.

Recent advancements in artificial vision and CNN models have led to various innovative approaches for fruit grading. For instance, Sun et al. [188] employed a combination of

Support Vector Machines (SVM), Deep Convolutional Generative Adversarial Networks (DCGAN), and CNN models to classify Red Fuji apples by color, size, and defects, achieving an accuracy of 96.50% using transfer learning on their segmented dataset of over 10,000 images. Similarly, Naik [131] proposed four CNN-based models to categorize mangoes into four quality classes by evaluating parameters such as ripeness, shape, and size, and benchmarked these models against traditional machine learning methods using the Mango-Sapan Naik dataset. Tripathi and Maktedar [200] introduced a CNN model to automate mango grading into three quality categories based on health, maturity, and size. The current literature clearly demonstrates that CNN-based image analysis for external fruit quality parameters such as color and overall appearance achieves higher accuracy compared to traditional computer vision and machine learning techniques.

Several previous reviews have examined computer vision applications for assessing the external quality of fruits. For example, Liu et al. [115] reviewed traditional machine learning techniques and DL models for fruit image classification, highlighting the limited but promising application of CNNs at that time and recommending their increased use in future studies. Similarly, Hameed et al. [65] compared various computer vision techniques for fruit and vegetable quality assessment, although their review cited only a few studies employing CNNs. Naranjo-Torres et al. [132] analyzed non-destructive techniques for fruit quality inspection, extensively reviewing CNN-based methods from over 40 studies, categorizing them into classification, quality control, and automatic harvesting detection. Bhargava and Bansal [19] discussed image-processing methods for fruit inspection, emphasizing CNN models' potential to significantly improve classification accuracy based on color, size, texture, shape, and defects. Prabhu et al. [163] explored diverse computer vision and preprocessing techniques aligned with automation in fruit quality inspection, but mentioned CNN models only briefly. Finally, Aherwadi and Mittal [2] reviewed traditional and CNN-based approaches for fruit image classification and ripeness detection, noting CNNs' superior performance metrics compared to traditional methods but highlighting gaps regarding fruit shapes, defects, and the reliability of certain reviewed articles.

As previously mentioned, researchers have increasingly adopted CNN models, although some studies continue to utilize traditional computer vision and machine learning methods. While Naranjo-Torres et al. [132] presented a comprehensive review of CNN applications covering both harvest and post-harvest stages, the present study specifically emphasizes the post-harvest phase, as it is crucial for categorization and commercialization. Most of the reviewed literature does not address this particular stage comprehensively, nor does it thoroughly evaluate the robustness of datasets used for training and testing CNN models. Additionally, evaluations typically lack comparative analyzes across multiple CNN architectures. Therefore, this work explicitly addresses these gaps by examining the robustness of datasets and comparing various CNN models specifically applied to the external quality inspection of fruits.

Given the growing adoption and effectiveness of CNN models, this work reviews recent advances in CNN-based approaches for external quality assessment of fruits through image analysis of parameters such as ripeness, deformation, and defects. Compared to previous reviews, this paper contributes uniquely by providing a detailed analysis of CNN model applications specifically for post-harvest fruit classification aligned with international standards. It examines the most relevant literature from 2018 to 2023, summarizing essential details, including datasets, data types, inspection parameters, CNN architectures, and their reported performance results.

As the use of CNN models for fruit quality control has surged in recent years, traditional computer vision and conventional machine learning techniques remain recognized for their robustness and effectiveness. This diversity of methodologies highlights the continued value of established approaches alongside CNN-driven advances. To evaluate the scientific landscape, we conducted a preliminary analysis of CNN applications in external fruit inspection, focusing on publication volume, model architectures, and evaluation metrics. Figure 2.3 shows that most CNN-based studies appeared in 2021, while recent years have seen improvements in accuracy through more robust network designs. Notably, research on ripeness, deformation, and defect parameters peaked in 2021, whereas defect-focused publications have maintained steady output, indicating opportunities for further innovation in multi-parameter assessments.

Figure 2.4 illustrates the relative prevalence of CNN architectures in external fruit quality inspection. ResNet50 and VGG16 emerge as the most widely adopted models, offering a favorable balance between network depth and computational efficiency. InceptionV3 and VGG19 follow closely, leveraging advanced feature extraction, while AlexNet and MobileNet appear as lighter-weight alternatives that maintain practical accuracy in resource-constrained scenarios.

2.3 Data Acquisition

The adoption of machine learning has become essential across industries, including agriculture, where external fruit quality inspection can be greatly enhanced by Transformer and CNN models trained on extensive, high-quality image datasets [129]. Optimal model performance relies not only on dataset size but also on the rigor of image acquisition, which must account for environmental variables such as lighting, background reflectance, color camouflage, and

Fig. 2.3 Annual quantity of studies employing CNN models for fruit classification, considering ripeness, deformities, and defects parameters.

object occlusion [65, 132, 163]. Careful control of these factors during data collection yields robust and reliable datasets, thereby improving the accuracy and generalizability of DL methods in post-harvest fruit assessment.

High-quality images with precise annotations of the target parameters are critical for constructing a reliable fruit image dataset. Acquiring samples across diverse environments and lighting conditions enhances model robustness and generalizability. Building a large, balanced dataset that fully represents each quality parameter augmented with synthetic images as needed and applying rigorous cleaning to remove duplicates and low-quality captures ensures the development of accurate CNN models for fruit classification in accordance with international standards.

2.3.1 Types of Images according to the Spectrum

In the context of computer vision, an image is a graphical representation of an object, scene, or concept composed of discrete elements called pixels. Its resolution governs the level of detail, while the file format determines how visual data is encoded and accessed. Color characteristics, luminous intensity, and the spatial distribution of pixels further enrich the information content, supporting accurate interpretation. Reliable image acquisition and

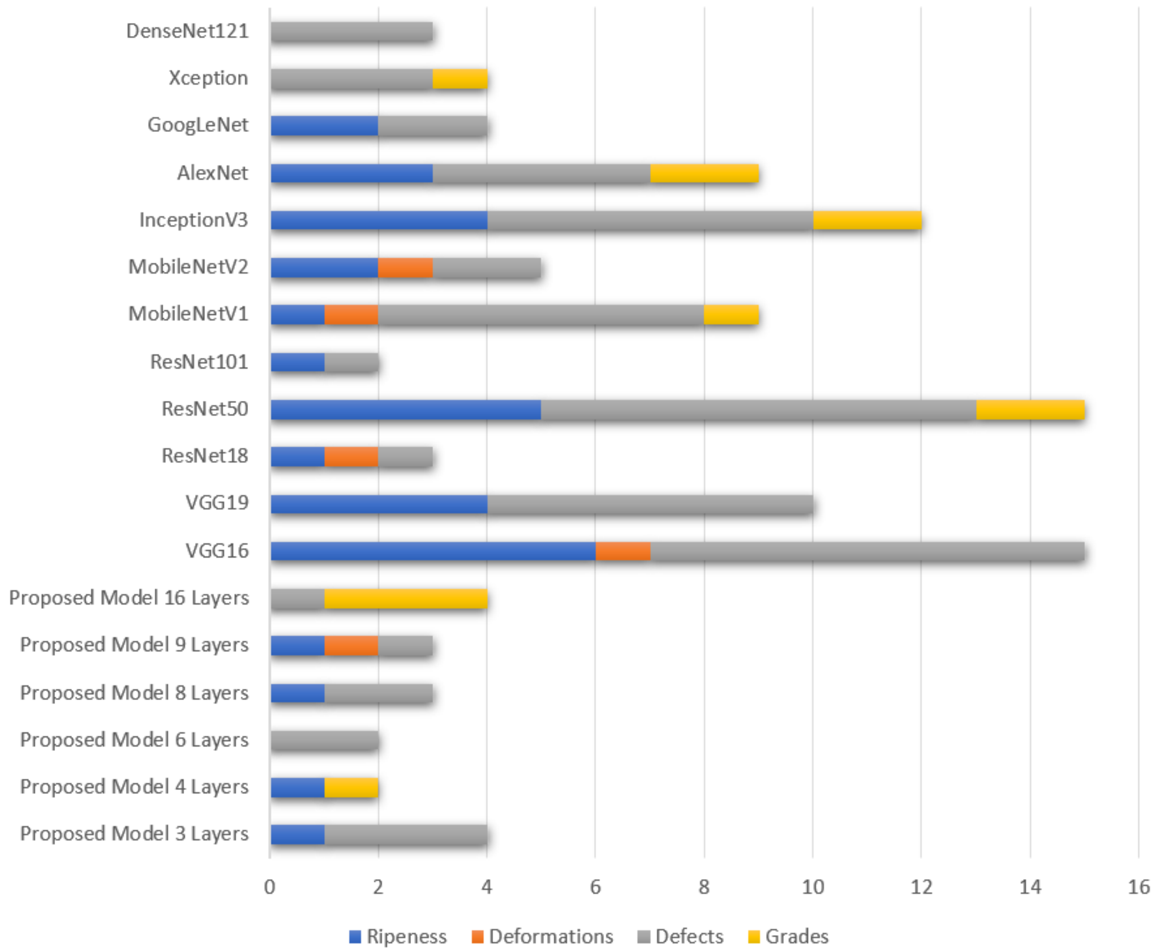


Fig. 2.4 CNN models utilized for identifying external quality inspection parameters in fruits.

processing are fundamental to applications ranging from photography and graphic design to advanced artificial vision systems.

Figure 2.5 illustrates how visible and multispectral imaging extend data acquisition beyond human vision. Visible imaging reproduces wavelengths perceptible to the human eye, providing accurate representations of external appearance. Multispectral techniques capture discrete bands such as visible (Vis) and near-infrared (NIR) to highlight physiological and structural features relevant to precision agriculture and remote sensing [38]. Employing these modalities enables consistent adherence to international quality control standards, thereby enhancing the precision and reliability of post-harvest fruit assessments.

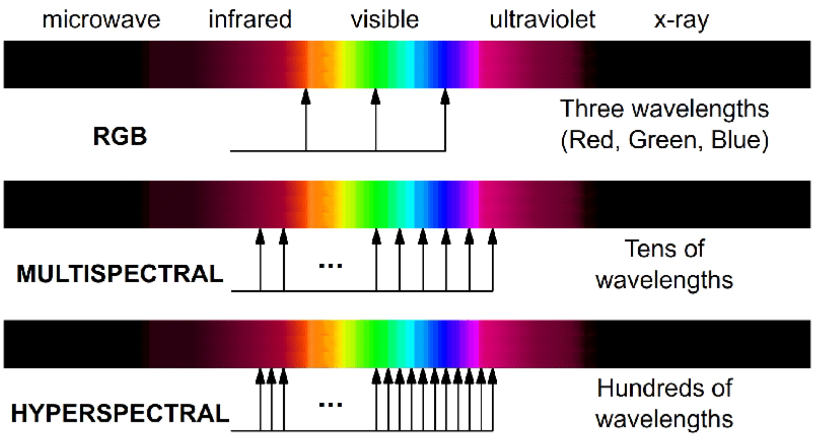


Fig. 2.5 Electromagnetic Spectrum [220].

Image processing in the visible spectrum offers a cost-effective approach for fruit classification using CNN models. Conventional RGB (red, green, and blue) cameras capture three monochrome channels at 700.0 nm (red), 546.1 nm (green) and 435.8 nm (blue), closely emulating human vision to evaluate external attributes such as ripeness, deformation and defects, albeit with limited spectral range [207, 223]. Beyond the visible range, infrared video sensors have achieved real-time lesion detection on melon skin with 97.5% accuracy [193]. Near-infrared spectroscopy, probing 780–2,500 nm , acquires reflectance and transmittance spectra that reflect fruit chemical composition and structural characteristics [138]. Moreover, Vis–NIR imaging across 300–1,000 nm combined with CNN classifiers has yielded 85% accuracy in apple bruise detection [52]. These modalities underscore the balance between cost, spectral richness and classification performance essential for efficient post-harvest quality inspection.

Multispectral imaging systems capture two or more monochrome images across distinct spectral bands, from visible to infrared and ultraviolet, enabling detection of features imperceptible to the human eye [164, 207, 223]. This method permits inspection in real time with fewer filters than hyperspectral approaches [202], but it requires careful calibration to prevent image misalignment and distortion [121]. Although research on CNN models for multispectral fruit imaging is less extensive than for the visible spectrum, early results show promise for evaluating ripeness, firmness, sugar content and surface defects. Ongoing advances in sensor technology and image processing algorithms will reduce cost and complexity, broadening the use of multispectral imaging for efficient quality control after harvest.

Upon concluding this section, it is important to clarify that this doctoral thesis will focus on external fruit quality assessment using visible RGB imaging. Although multispectral techniques offer additional insights, the limited availability of multispectral datasets restricts their current application. Future research should therefore prioritize the development and sharing of comprehensive multispectral fruit image repositories to enable deeper exploration of these advanced quality-control methods.

2.3.2 Real or Synthetic Data

This section emphasizes the critical importance of comprehensive reference datasets in fruit quality classification. High quality image repositories are essential for training, testing and validating DL models, as they enable meaningful performance comparisons and robust evaluation. Despite their value, publicly accessible datasets remain scarce due to the considerable time and expense required to collect, annotate and curate large scale fruit image collections.

Real image datasets, acquired with various sensor technologies [65], are often compiled in controlled settings that address device characteristics, illumination variability, background occlusion, temperature fluctuations, and framing considerations [163]. Nevertheless, publicly accessible repositories seldom offer the volume and diversity required for effective CNN training and evaluation. To bridge this gap, the synthesis of artificial images has become a practical approach to augment real datasets, enhancing scale, diversity, and ultimately model robustness and generalization.

Several studies have released publicly accessible real-world fruit image datasets targeting external quality parameters such as ripeness, deformations and surface defects. Table 2.2 summarizes these resources by fruit species, image volume and annotation criteria, providing a foundation for standardized evaluation and comparative analysis of inspection models.

Table 2.2 Summary of Real Fruit Datasets with Public Access

Articles	Fruits	Dataset	Parameters
[90]	Apples, Bananas, Oranges	https://www.kaggle.com/datasets/sriramr/ fruits-fresh-and-rotten-for-classification	Ripeness, Defects
[132]	Apples	https://www.kaggle.com/moltean/fruits	Ripeness, Defor-mities, Defects
-	Apples	https://www.kaggle.com/datasets/ projectlzp201910094/applescabfds	Defects
-	Apples, Bananas, Oranges	https://zenodo.org/record/4788775# .ZBpLXHaZNPY	Ripeness, Defects
-	Apples, Avocados, Bananas, Blackberries, Cherries, Grapes, Kiwis, Lemons, Mangoes, Oranges	https://zenodo.org/record/4788775# .ZBpLXHaZNPY	Fresh
[79]	Apples	https://github.com/AsIsmail/Apple-Dataset	Ripeness, Defor-mities, Defects
-	Bananas, Lemons, Mangoes, Oranges, Strawberries	https://data.mendeley.com/datasets/6ps7gtp2wg/1	Defects

Table 2.2 Summary of Public Fruit Datasets (continued)

Articles	Fruits	Dataset	Parameters
[131]	Mangoes	https://data.mendeley.com/datasets/nsjggt7tyz/1	Ripeness, Deformities, Defects
[200]	Mangoes	https://data.mendeley.com/datasets/fmfncxjz3v/1#folder-2283e5f3-1ee9-44cc-a200-aaa079161167	Ripeness, Deformities, Defects
-	Mangoes	https://www.kaggle.com/datasets/saurabhshahane/mango-varieties-classification	Ripeness, Deformities, Defects
-	Mangoes	https://www.kaggle.com/datasets/mypapit/sala-mango-image	Defects
[22]	Lemons	https://github.com/jordan-bird/synthetic-fruit-image-generator	Deformities, Defects
[34]	Bananas	https://github.com/luischuquim/BananaRipeness	Ripeness
[147]	Apples, Mangoes	https://github.com/luischuquim/Healthy-Defective-Fruits	Ripeness
[17]	Apples, Mangoes, Strawberries	https://github.com/luischuquim/Deformed-fruits	Deformities

To overcome the scarcity and variability of real fruit images, researchers have turned to synthetic datasets that facilitate precise control over lighting, background and defect parameters while enabling rapid generation of large volumes of labeled data [180, 207]. By replicating key visual cues such as color transitions, surface deformations and lesion patterns, these synthetic images provide a viable alternative for training and validating CNN models in external quality inspection, thereby enhancing robustness and generalization when real datasets are limited.

Current image synthesis platforms such as Unreal Engine, DALL·E Mini and Stable Diffusion enable the rapid generation of realistic fruit images to augment real datasets. By employing transfer learning, these synthetic samples pre-train convolutional networks, providing robust initial weights that reduce dependence on extensive real-data annotations and accelerate convergence during fine-tuning. This strategy optimizes both time and resources in the development of computational intelligence methods [17, 34, 147].

2.4 Fruit Quality Parameters

Fruit quality assessment for commercial grading is commonly structured around external, image-observable attributes that directly influence consumer perception and market value. In this section, the reviewed literature is organized under the general concept of Fruit Quality Parameters, which encompasses three parameters that can be reliably estimated from visual information: ripeness, which reflects the physiological state of the fruit and is primarily manifested through variations in color and texture; deformation, which describes deviations in size and shape that affect uniformity and classification; and defects, which include surface or subsurface alterations associated with mechanical damage, physiological stress, or pathological deterioration. Recent advances in convolutional neural networks have enabled robust learning of visual features for these attributes, while accounting for dataset scale, image acquisition conditions, and alignment with international grading standards.

2.4.1 Ripeness Parameter

Ripeness critically influences consumer choice in retail environments, with color serving as the dominant visual cue for maturity assessment [18, 19, 65, 223]. Recent work has leveraged convolutional neural networks to automate ripeness classification, achieving high accuracy across a range of fruit species. The following section reviews these state-of-the-art CNN-based approaches to maturity estimation.

Chuquimarca et al. [34] address banana ripeness estimation through a hybrid strategy that combines real and synthetic images across multiple maturity levels. The study proposes a simple CNN architecture enhanced via transfer learning and evaluates multiple architectures, hyperparameter settings, and optimizers, achieving a peak accuracy of 91.70%. A key contribution is demonstrating that synthetic data generation can substantially reduce acquisition cost and time; however, the lack of balance among real images per ripeness class introduces potential bias and may overestimate generalization. Consequently, class rebalancing and stratified sampling are essential methodological steps for future extensions.

Along similar lines, Aherwadi et al. [3] evaluate a custom CNN and AlexNet on both an in-house dataset and the public Fruit-360 dataset, reporting up to 99.44% accuracy on augmented in-house data but only 81.96% on Fruit-360. This marked performance gap indicates strong domain sensitivity and underscores the importance of cross-dataset validation and robustness analysis. Although the study surveys 35 related works, its comparative contribution is weakened by the absence of explicitly standardized ripeness definitions aligned with international grading schemes and by the lack of direct benchmarking against prior methods on equivalent datasets.

Zhu and Spachos [227] combine SVM and YOLOv3 on a dataset of 1,000 images generated from 150 originals through data augmentation, classifying two ripeness stages based on black spot counts with 90.16% accuracy. While the approach pragmatically links a visible attribute to ripeness, it remains limited to a reduced number of classes that do not reflect common commercial grading standards. Moreover, the small number of original samples and reliance on augmentation may constrain the robustness of deep detectors such as YOLOv3.

Focusing on computational efficiency, Saragih and Emanuel [176] compare MobileNetV2 and NASNetMobile using 436 images distributed across four ripeness classes, achieving 96.18% accuracy with faster inference for MobileNetV2. Despite these promising results, the limited dataset size and narrow architectural scope prevent definitive conclusions regarding model superiority, positioning the findings as preliminary evidence for lightweight deployments rather than as a comprehensive evaluation.

At the level of transfer learning with scarce data, Rivero Mesa and Chiang [171] fine-tune VGG16 to classify banana cluster quality and report satisfactory accuracy with minimal samples. Nevertheless, the lack of systematic comparison with baseline models, limited dataset expansion, and absence of publicly released trained models reduce reproducibility and make it difficult to assess performance gains relative to modern architectures.

Ramadhan et al. [167] integrate a YOLO-based segmentation stage prior to VGG16-based classification of four banana ripeness levels, observing marginal differences between SGD (94.12%) and Adam (93.25%) on a dataset of 300 images. While the segmentation–classification pipeline is relevant for real-world scenarios with background clutter, the small dataset and reliance on a single backbone limit the interpretability of the observed gains, highlighting the need for broader architectural comparisons and additional evaluation metrics.

Addressing finer-grained maturity levels, Ni et al. [137] apply GoogLeNet with transfer learning to classify six banana ripeness stages, achieving 98.92% accuracy, and further demonstrate cross-fruit generalization to strawberries with 92.47%. Although these results suggest strong transferability, the study lacks a detailed description of ripeness labeling criteria and does not discuss dataset bias, class imbalance, or overfitting, all of which are critical when reporting near-ceiling accuracies.

With a larger-scale dataset, Zhang et al. [224] train a CNN on 17,312 banana images spanning seven maturity classes and achieve 95.60% precision, outperforming traditional machine learning approaches. The dataset scale strengthens training reliability; however, the analysis is confined to a single architecture without optimizer tuning or comparison to contemporary CNNs, limiting insight into whether performance gains stem primarily from data volume or model design.

For mango ripeness, Kumari et al. [103] propose a self-adaptive chicken swarm optimization strategy integrated with a hybrid fuzzy–CNN model, attaining 91.75% accuracy on the Kesar mango dataset. While the use of metaheuristic optimization is innovative, the overall grading pipeline still relies on traditional techniques for several stages, fragmenting the end-to-end learning process and leaving room for improvement through fully deep learning–based frameworks.

Similarly, Dutta et al. [49] evaluate ResNet50, InceptionV3, and VGG16 on 826 mango images across four ripeness classes, identifying VGG16 as the best-performing model with 96.2% accuracy and extending the analysis to shelf-life prediction via support vector regression. Although this integration broadens postharvest decision support, the limited dataset size constrains generalization, even with augmentation, indicating the need for larger-scale validation.

In less-studied fruits, Azadnia et al. [11] acquire hawthorn images under controlled illumination and compare InceptionV3, ResNet50, and a custom CNN, reporting 100% validation accuracy for InceptionV3. While controlled acquisition enhances consistency, perfect validation performance on small datasets raises concerns of overfitting or lenient splits, and the absence of explicit grading standards limits applicability to real commercial settings.

Exploring spectral diversity, Garillos-Manliguez and Chiang [53] investigate multimodal CNNs that fuse RGB and hyperspectral data for six papaya ripeness stages, achieving 88.64% accuracy with MD-VGG16. However, unimodal configurations sometimes match or exceed multimodal performance, suggesting suboptimal fusion strategies and highlighting the need to carefully assess the added value and cost of hyperspectral sensing.

In a related study, Behera et al. [16] apply transfer learning with seven CNN architectures to a small papaya dataset of 300 images and report 100% accuracy with VGG19. Despite the impressive result, the limited dataset size and controlled acquisition conditions increase the risk of overfitting, reducing confidence in real-world robustness without more challenging validation protocols.

For mangosteen, Mohtar et al. [127] train an InceptionV3-based CNN on 800 images across six maturity levels, achieving 91.90% accuracy along with balanced precision, recall, and F1 scores. While the inclusion of multiple metrics strengthens evaluation, the modest dataset size and lack of extensive benchmarking restrict conclusions regarding competitiveness with state-of-the-art models.

In the context of orange ripeness, Asriny et al. [9] train a CNN on 1,000 smartphone-acquired images and report 96% accuracy using ReLU and 93.80% using Tanh activation. Nevertheless, insufficient detail on image acquisition protocols, background conditions, and alignment with grading standards undermines reproducibility and limits fair comparison with other studies.

For grape ripeness, Ramos et al. [168] compare a custom CNN with VGG19 on Syrah and Cabernet Sauvignon varieties, achieving 93.41% and 72.66% accuracy, respectively. While the methodological description is thorough, the limited varietal scope and performance disparity across cultivars suggest strong cultivar dependency, calling for broader validation across grape types and environments.

Adopting a multi-fruit perspective, Sumathi and Vinod [187] train MiniVGGNet with SGD on 6,702 images of oranges, papayas, and bananas, achieving 96.56% accuracy. However, the absence of comparative benchmarking against standard architectures or prior studies restricts assessment of model superiority, rendering the result primarily an internal reference.

Likewise, Pardede et al. [153] modify VGG16 by replacing the top layers with a multi-layer perceptron and apply various regularization strategies to classify two ripeness states across multiple fruits, achieving precisions between 84% and 90%. Although the focus on regularization is methodologically sound, missing details regarding dataset size and composition hinder evaluation of robustness and stability.

Finally, Rodriguez et al. [172] propose a nine-layer CNN with data augmentation to classify three ripeness states across apples, bananas, strawberries, and oranges, reporting accuracies ranging from 82.60% to 99.12%. While the multi-fruit approach is practically relevant, limited dataset transparency and heavy reliance on augmentation increase the risk of overfitting, emphasizing the need for clearer dataset documentation and external validation to support the reported performance.

Table 2.3 summarizes current CNN-based ripeness detection efforts, highlighting bananas as the most studied fruit with a wider range of maturity classes than others. Ripeness classes vary across species and even within the same fruit, indicating the need for more consistent classification schemes. Grapes remain underrepresented, offering a clear avenue for future investigation. Most studies use datasets of 1,000 to 10,000 images, emphasizing dataset size as a critical factor in model training and evaluation. The predominance of real RGB imagery also suggests an opportunity to explore synthetic and hyperspectral data to enhance model robustness and accuracy.

Table 2.3 Comparative summary of ripeness dataset parameters

Articles	**Fruits**	**Dataset size (images)**	**Data type / Ripeness levels**
[34]	Banana	Medium (1,000–10,000)	RGB real / 4
[3]	Banana	Medium (1,000–10,000)	RGB real / 3
[227]	Banana	Small (<1,000)	RGB real / 5
[176]	Banana	Small (<1,000)	RGB real / 4
[171]	Banana	Small (<1,000)	RGB real / 4
[167]	Banana	Small (<1,000)	RGB real / 4
[137]	Banana	Medium (1,000–10,000)	RGB real / 6
[224]	Banana	Large (>10,000)	RGB real / 7
[103]	Mango	Medium (1,000–10,000)	RGB real / 4
[49]	Mango	Medium (1,000–10,000)	RGB real / 4
[11]	Hawthorn	Medium (1,000–10,000)	RGB real / 3
[53]	Papaya	Medium (1,000–10,000)	RGB real, multispectral / 6
[16]	Papaya	Small (<1,000)	RGB real / 3
[127]	Mangosteen	Small (<1,000)	RGB real / 6
[9]	Orange	Medium (1,000–10,000)	RGB real / 5
[168]	Grape	Medium (1,000–10,000)	RGB real / 3–4
[187]	Banana, Papaya, Orange	Medium (1,000–10,000)	RGB real / 3
[153]	Mango, Apple, Orange	Medium (1,000–10,000)	RGB real / 2
[172]	Apple, Orange, Strawberry	Small (<1,000)	RGB real / 3

Table 2.4 summarizes CNN architectures applied to fruit ripeness classification. Figure 2.6 illustrates the relative frequency of each model, with VGG16 most prevalent, followed by ResNet50, VGG19 and InceptionV3. Research has focused on maturity estimation in bananas, mangoes, papayas and oranges, employing CNNs enhanced by transfer learning and data augmentation to achieve high accuracy. Common limitations include limited dataset sizes, narrow architectural comparisons and misalignment of ripeness categories with international grading standards.

Table 2.4 Comparative summary of CNN models for fruit ripeness classification

Articles	CNN Layers	VGG Models	Other CNN Models
[11]	8	–	ResNet-50, InceptionV3
[34]	3 (CIDIS)	VGG-19	ResNet-50, Inception-ResNetV2, InceptionV3
[3]	5	–	AlexNet
[103]	SA-CSO-FCNN	–	–
[187]	–	VGG-16	–
[16]	–	VGG-16, VGG-19	ResNet-18, ResNet-50, ResNet-101, AlexNet, GoogLeNet
[49]	–	VGG-16	ResNet-50, InceptionV3
[53]	–	VGG-16, VGG-19	ResNet-50, ResNet-x50, MobileNetV1, MobileNetV2, AlexNet
[153]	–	VGG-16	–
[168]	10	VGG-19	–
[171]	–	VGG-16	–
[172]	9	–	–
[176]	–	–	MobileNetV2, MobileNet-NAS
[227]	–	–	YOLOv3
[9]	3	–	–
[137]	–	–	GoogLeNet
[167]	–	VGG-16	–
[127]	–	–	InceptionV3
[224]	4	–	–

Reviewed studies address ripeness classification and prediction for bananas and other fruits such as mangoes, papayas, hawthorns, and oranges. They rely on convolutional neural networks enhanced with transfer learning and data augmentation to improve generalization, reporting accuracies between 81.75% and 99.44% depending on the dataset and model.

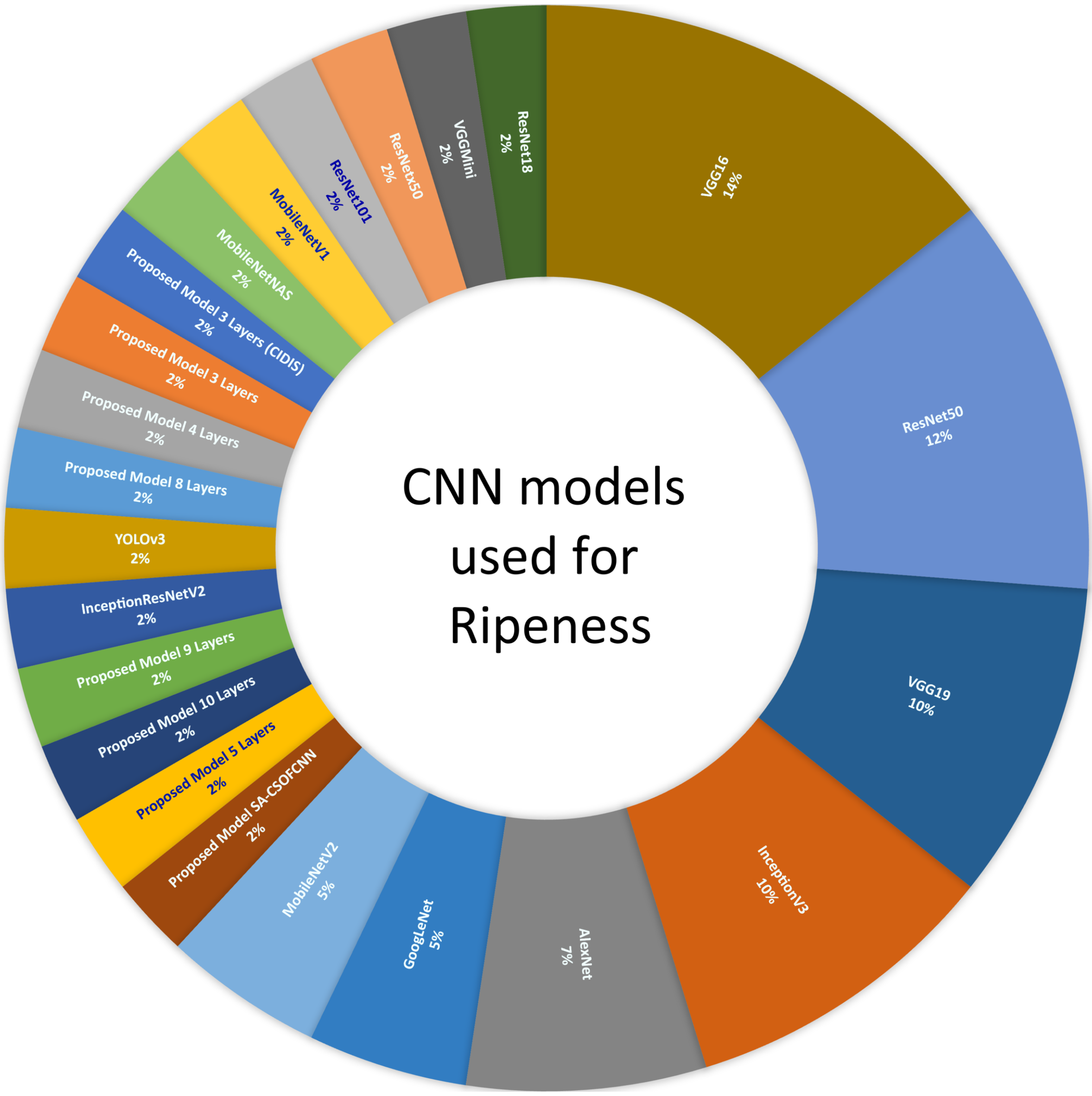


Fig. 2.6 The percentage distribution of CNN model usage for identifying fruit ripeness.

Common limitations include small dataset sizes, limited benchmarking against diverse baselines, and inconsistent ripeness labeling relative to international standards. Future research should expand and diversify datasets, evaluate a broader range of architectures, align ripeness categories with standardized criteria, and implement robust validation protocols to identify and mitigate bias and overfitting.

2.4.2 Deformation Parameter

Size and shape are fundamental geometric attributes in external fruit quality inspection, directly influencing classification and market value [134]. Shape is quantified through contour analysis and interior features, incorporating spectral properties, curvature, and structural descriptors [17]. Key metrics such as curvature, corner points, regional segmentation, centroid, convexity, circularity, and eccentricity jointly characterize fruit morphology [65].

Cao et al. [27] propose LightNet, a dedicated CNN architecture for automatic apple grading based on size, achieving 99.31% accuracy on a dataset of 6,669 images distributed across three diameter classes and outperforming established architectures such as AlexNet, VGG11, ResNet18, MobileNet variants, and ShuffleNetV2. A notable contribution of this work is the cross-domain evaluation strategy, where the model is trained on zizania images and tested on apple images, yielding comparable performance and thus suggesting a degree of morphological generalization. While this result strengthens the claim of adaptability, the analysis remains focused on diameter-based grading under controlled conditions, and further validation on more diverse fruit shapes, acquisition environments, and independent datasets would be required to fully substantiate robustness in real commercial scenarios.

In contrast, Rivero Mesa and Chiang [171] adapt the VGG16 architecture for non-invasive banana size classification and report satisfactory accuracy despite relying on a limited number of training images. The study highlights the potential of transfer learning to mitigate data scarcity; however, the absence of systematic benchmarking against alternative CNN architectures restricts assessment of the proposed model's relative effectiveness. The authors themselves acknowledge the need for larger datasets or data augmentation to reduce overfitting, which suggests that the reported performance may be sensitive to sample size. Moreover, although the image dataset is described as open access, the lack of availability of trained models limits reproducibility and constrains further comparative evaluation by the research community.

Focusing on fruit shape as a quality parameter, Momeny et al. [128] introduce a CNN-based approach for cherry shape classification and demonstrate clear superiority over traditional feature-based methods, including Histogram of Oriented Gradients, Local Binary Patterns, K-Nearest Neighbors, Artificial Neural Networks, fuzzy logic, and ensemble decision trees, achieving 99.40% accuracy. The dataset initially comprises 719 RGB images and is expanded to 14,380 samples through data augmentation, which substantially improves training stability. Nevertheless, the evaluation is restricted to comparisons with classical techniques, and the absence of benchmarking against widely used CNN architectures limits insight into the proposed model's competitiveness within the broader deep learning landscape. As such, while the results confirm the effectiveness of CNNs for geometric quality assessment, broader architectural comparisons would strengthen the generalizability of the conclusions.

Table 2.5 summarizes research on fruit shape and size identification, with apples as the most studied species followed by bananas. Studies addressing deformities are limited, creating an opportunity to close this gap and support the standardization of external quality parameters for commercial grading. Deformities are defined with two or three levels in the surveyed works, and dataset sizes vary widely; however, reliable CNN training typically requires more than 1,000 images per class. All reviewed approaches use RGB imagery for these geometric assessments.

Table 2.5 Comparative summary of deformation dataset parameters

Articles	Fruits	Dataset size (images)	Data type / # classes
[27]	Apple	Medium (1,000–10,000)	RGB real / 3
[171]	Banana	Small (<1,000)	RGB real / 3
[128]	Cherry	Large (>10,000)	RGB real and synthetic / 2

Table 2.6 summarizes the CNN architectures employed for fruit deformation classification. Figure 2.7 shows their relative usage, revealing that several models are adopted with comparable frequency. Reported accuracies span from 92.98% to 99.04%. Notably, architectures such as VGG16 and ResNet50 are shared with ripeness classification, indicating their effectiveness across related external quality assessment tasks.

Table 2.6 Comparative summary of CNN models for fruit deformation classification

Articles	CNN Layers (Proposed)	VGG Models	ResNet Models	Other CNN Models
[27]	-	VGG-11	ResNet-18	MobileNet-V1, MobileNet-V2, ShuffleNetV2, LightNet
[171]	-	VGG-16	-	-
[128]	9	-	-	-

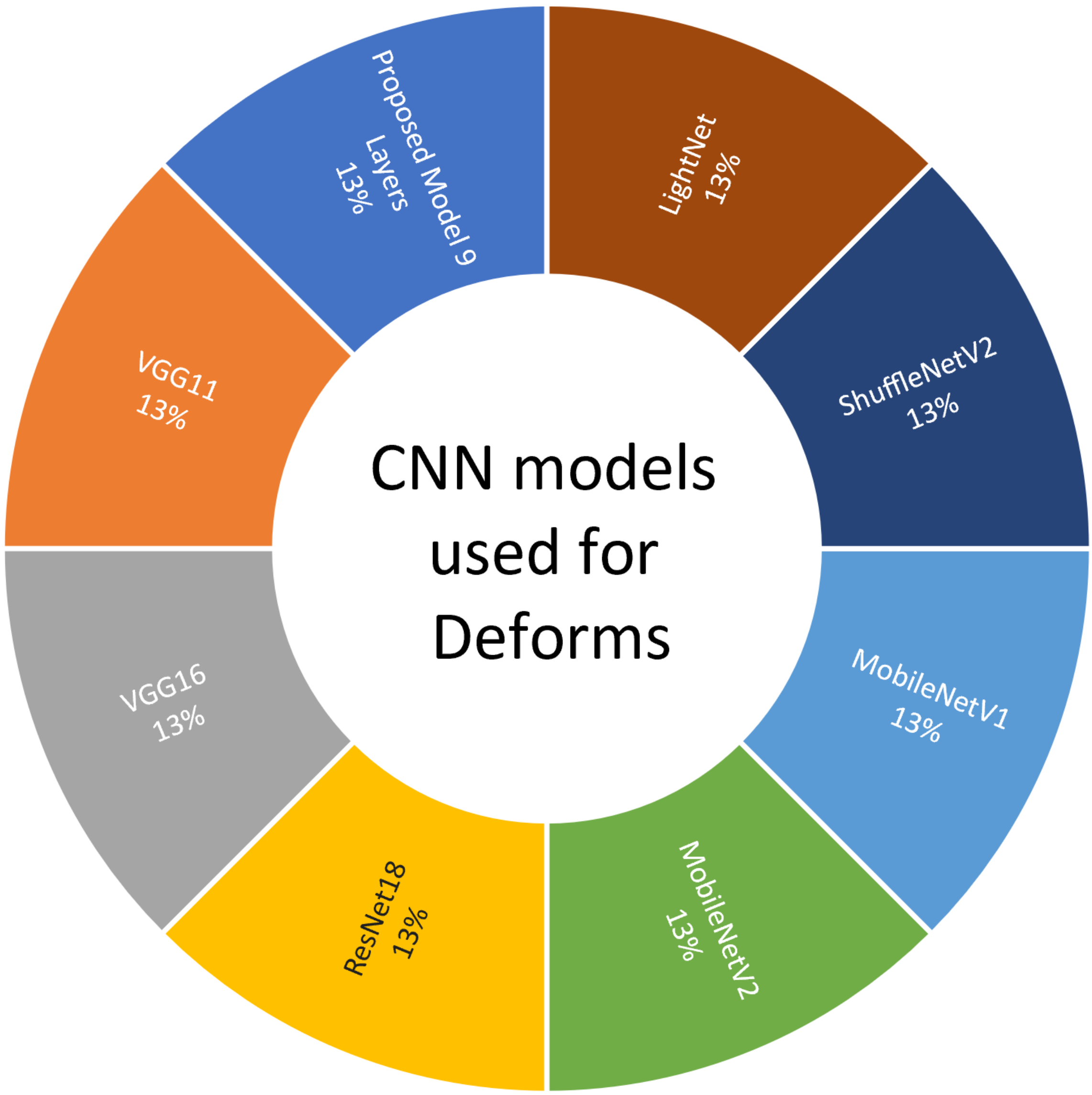


Fig. 2.7 The percentage distribution of CNN model usage for identifying fruit deformations.

Reviewed studies show that convolutional neural networks accurately classify fruits such as apples, bananas, and cherries using shape and size features, outperforming traditional methods. Establishing the consistent superiority of new models requires broader comparative evaluations against publicly available baselines. Increasing dataset sizes and sharing trained models are also necessary to improve reproducibility and accelerate progress in this domain.

2.4.3 Defect Parameter

Fruit defects stem from pre-harvest practices, post-harvest handling and storage conditions and encompass a range of disorders [203]. Pathological damage arises from microbial pathogens, while mechanical injuries caused by compression or impact lead to tissue failure, pigment breakdown and increased susceptibility to infection [182]. Physiological stress, influenced by nutrition, temperature and respiration, can manifest as bitterness, browning or internal drying [169]. Morphological deformations alter fruit shape and complicate vision-based detection despite preserving compositional integrity, and latent internal defects may result from any of these factors [140]. These surface and subsurface anomalies critically affect external quality inspection, often leading to rejection, and their reliable detection depends on the imaging modality, visible or multispectral, and on controlled lighting, scale, and viewpoint during acquisition [223].

Buyukarikan and Ulker [25] investigate non-destructive detection of physiological disorders in apples using CNNs trained on images acquired under varying lighting conditions, camera angles, and distances. Among the evaluated architectures, Xception achieves the highest precision, recall, and F1 score, supported by statistical significance tests. While the experimental design strengthens robustness against acquisition variability, the exclusive focus on physiological disorders excludes other common defect types, thereby limiting applicability to comprehensive quality inspection. Moreover, the notably poor performance of VGG19 is not sufficiently analyzed, leaving open questions regarding the influence of model complexity and parameterization on generalization.

In a more constrained setting, Agarwal et al. [1] address the detection of scar, scab, and rot in apples using a shallow three-layer CNN trained on an augmented dataset of 500 images. The reported accuracy of 96.87% suggests that lightweight architectures can be effective for specific defect categories. However, the evaluation relies solely on accuracy, omitting precision, recall, and F1 score, which restricts interpretability, particularly in imbalanced scenarios. Furthermore, the acknowledgment of a Random Forest classifier achieving 99.00% accuracy without direct comparison weakens the argument for the CNN's superiority.

Focusing on binary quality assessment, Ibrahim et al. [77] compare a conventional CNN with AlexNet and GoogLeNet for fresh versus rotten apple classification using publicly available Kaggle images and standard augmentation techniques. The perfect accuracy achieved by GoogLeNet highlights the effectiveness of deep pre-trained models in capturing discriminative defect features, whereas the significantly lower performance of the conventional CNN underscores the limitations of shallow architectures. Nevertheless, the absence of cross-dataset validation raises concerns regarding potential overfitting to the selected benchmark.

Arango et al. [6] evaluate VGG19, ResNet50, and DenseNet121 for apple bruise detection using both RGB and near-infrared images captured under controlled illumination. DenseNet121 performs best on RGB data, while ResNet50 excels on NIR images, indicating that optimal architectures may depend on the spectral modality. Despite these insights, the small dataset size limits robustness, and the lack of RGB–NIR fusion prevents full exploitation of complementary spectral information that could enhance defect discrimination.

Addressing multi-grade quality classification, Li et al. [111] propose a proprietary CNN to categorize apple surface defects into premium, middle, and poor classes, achieving 95.33% accuracy on a 300-image dataset and outperforming InceptionV3 and HOG/GLCM+SVM pipelines. Although the results indicate the promise of task-specific CNN designs, reliance on accuracy alone and limited dataset scale weaken the strength of the conclusions. A more comprehensive metric set and deeper analysis of competing methods would be required to substantiate practical relevance.

In a broader benchmarking effort, Naranjo-Torres et al. [132] compare AlexNet, VGG16, MobileNet, InceptionV3, and ResNet50 on the Apple-NDDA dataset comprising 1,110 images. VGG16 emerges as the most accurate model at 89.58%, illustrating the competitiveness of relatively deep but well-established architectures. However, the moderate dataset size and lack of optimizer tuning suggest that the reported performance may not reflect the full potential of the evaluated models.

Similarly constrained by data volume, Wu et al. [215] modify AlexNet to an eleven-layer configuration for apple defect detection while explicitly distinguishing stems and calyx to reduce false positives. The proposed CNN achieves 86% accuracy, outperforming classical baselines such as BP, PSO, and SVM. Despite this improvement, the small dataset of 500 samples limits generalization, indicating that the gains may not persist under more diverse conditions.

In contrast, Fan et al. [51] exploit a large-scale dataset of 79,200 RGB images to train an eight-layer CNN for detecting multiple apple defects, including insect damage, rot, and scarring. The reported accuracy of 90.90% benefits from the substantial data volume, reinforcing the importance of scale for robust training. Nonetheless, the absence of comparisons with modern CNN architectures and multispectral approaches restricts assessment of the model's competitiveness.

Extending defect detection across fruits, Pacheco et al. [147] evaluate InceptionV3, MobileNetV2, VGG16, and DenseNet121 on a mixed real and synthetic dataset of 20,000 apple and mango images. MobileNetV2 achieves the highest accuracy, demonstrating that lightweight architectures can outperform deeper models under certain conditions. However, limited details regarding acquisition protocols and lack of external benchmarking constrain conclusions about generalizability.

Robustness under domain shifts is explicitly analyzed by Tasara and Qomariyah [195], who apply transfer learning with VGG16, ResNet50, and MobileNetV3-small to classify multiple fruit types under degraded conditions such as blur and low light. VGG16 shows the highest resilience, yet all models suffer notable performance drops, emphasizing the vulnerability of CNNs to quality degradation. The study appropriately recommends expanding class definitions and incorporating low-quality samples during training to enhance robustness.

Addressing spoilage detection, Nazir et al. [133] compare SVM, ANN, and CNN models across apples, bananas, and oranges, reporting 96.85% accuracy for their CNN1 configuration. While the unified dataset facilitates comparison, the authors correctly identify dataset size and quality as limiting factors and propose multi-angle acquisition to improve coverage, highlighting practical considerations for real-world deployment.

In a binary assessment framework, Kumar et al. [101] introduce FruitCNN, achieving 99.60% accuracy on a private dataset of 12,000 images spanning six fruit types. Although the results surpass several public CNN baselines, the lack of explicit criteria defining optimal fruit conditions introduces ambiguity in label interpretation and may affect reproducibility.

A similar limitation appears in Pachón-Suescún et al. [148], who propose a DAG-based CNN to classify fresh versus spoiled fruits across eight varieties under varying acquisition conditions. The achieved accuracy of 94.43% demonstrates feasibility, yet the absence of benchmarking against established CNN architectures weakens claims of methodological advancement.

Using a public Kaggle dataset, Palakodati et al. [150] propose a custom CNN for fresh versus rotten fruit classification, outperforming transfer learning approaches with VGG16, VGG19, MobileNet, and Xception. While the analysis of hyperparameters is valuable, the limited dataset size restricts confidence in generalization beyond the benchmark.

Karakaya et al. [90] apply ResNet50 to classify fresh and rotten apples, bananas, and oranges, achieving 97.61% accuracy and outperforming classical vision methods. However, the small number of images per class and lack of comparison with other CNN architectures limit the strength of the conclusions.

Focusing on disease detection, Devi et al. [42] propose the TODN model, integrating orthogonal weight regularization with ThresholdedReLU activation within a DenseNet framework, and achieve 97.77% accuracy on the Fruit-360 strawberry dataset. The architectural innovation improves convergence, yet validation against a broader range of CNN baselines would better position the contribution within the literature.

In hyperspectral analysis, Pourdarbani et al. [162] employ 3D CNNs with spectral–spatial feature extraction to detect lemon bruises, achieving 90.47% accuracy with ResNet. The work effectively exploits spectral information and proposes adaptive learning-rate scheduling; however, the small dataset necessitates careful interpretation despite targeted augmentation strategies.

For mango surface defects, Nithya et al. [139] propose a CNN evaluated on 800 RGB images, achieving 98.50% accuracy and outperforming traditional methods. The lack of comparison with public CNN architectures, however, limits insight into relative performance.

Similarly, Duraisamy et al. [48] demonstrate that transfer learning significantly improves defect detection in mandarins using NIR images, with VGG16 accuracy increasing from 90% to 99.53%. While the results underscore the value of transfer learning and spectral imaging, extending the dataset to additional fruit varieties would strengthen claims of adaptability.

Investigating synthetic data, Bird et al. [22] train VGG16 on 3,000 synthetic lemon images and evaluate on real data, observing improved performance with synthetic pretraining. Despite this promising indication, broader architectural comparisons are needed to confirm whether the gains are model-specific or generally attributable to synthetic augmentation.

A hybrid approach is explored by Vasumathi and Kamarasan [205], who combine CNN feature extraction with LSTM classification for pomegranate defect detection, achieving high accuracy and balanced metrics on 6,519 images. The integration of temporal modeling enhances discrimination, although extension to multi-class disease detection remains necessary for practical deployment.

Perfect classification is reported by Jahanbakhshi et al. [81] using deep CNNs for sour lemon defect detection on 5,456 images. While impressive, the absence of comparisons with alternative CNN architectures limits confidence in the uniqueness of the reported performance.

Finally, Ashok and Vinod [8] apply MobileNetV1 to detect defects and chill damage in mangoes, achieving 96.40% accuracy on 2,130 images. The reliance on a single architecture and the merging of different defect types into one class constrain interpretability, suggesting that larger datasets and finer-grained labeling are required to validate the approach comprehensively.

Table 2.7 summarizes state of the art methods for fruit defect detection, with apples as the primary focus followed by bananas and oranges. These studies uniformly categorize samples as healthy or defective and employ datasets ranging from 1,000 to 10,000 images captured in RGB format. The consistency of dataset size and imaging modality highlights the need for expanding to larger and more varied collections and for evaluating additional spectral techniques to enhance defect detection performance.

Table 2.7 Comparative summary of defect dataset parameters

Articles	**Fruits**	**Dataset size (images)**	**Data type / Defect types**
[25]	Apple	Medium (1,000–10,000)	RGB real / 2
[1]	Apple	Medium (1,000–10,000)	RGB real / 3
[77]	Apple	Medium (1,000–10,000)	RGB real / 2
[6]	Apple	Medium (1,000–10,000)	RGB real, hyperspectral / 2
[111]	Apple	Small (<1,000)	RGB real / 2
[132]	Apple	Medium (1,000–10,000)	RGB real / 2
[215]	Apple	Medium (1,000–10,000)	RGB real / 3
[51]	Apple	Large (>10,000)	RGB real / 4
[147]	Mango, Apple	Large (>10,000)	RGB real, RGB synthetic / 3
[139]	Mango	Small (<1,000)	RGB real / 2
[8]	Mango	Medium (1,000–10,000)	RGB real / 2
[195]	Banana, Apple, Lima, Orange, Pomegranate	Medium (1,000–10,000)	RGB real / 2
[133]	Banana, Apple, Orange	Large (>10,000)	RGB real / 2
[101]	Banana, Apple, Lima, Orange, Guava, Pomegranate	Large (>10,000)	RGB real / 2

Table 2.7 Comparative summary of defect dataset parameters (continued)

Articles	Fruits	Dataset size (images)	Data type / Defect types
[148]	Banana, Mango, Orange, Strawberry, Lulo, Tamarillo, Tomato, Lemon	Medium (1,000–10,000)	RGB real / 2
[150]	Banana, Apple, Orange	Medium (1,000–10,000)	RGB real / 2
[90]	Banana, Apple, Orange	Medium (1,000–10,000)	RGB real / 2
[42]	Strawberry	Medium (1,000–10,000)	RGB real / 2
[162]	Lemon	Small (<1,000)	Hyperspectral / 3
[81]	Lemon	Medium (1,000–10,000)	RGB real / 2
[48]	Mandarin	Medium (1,000–10,000)	RGB real / 4
[23]	Tomato	Small (<1,000)	RGB real / 2
[205]	Pomegranate	Medium (1,000–10,000)	RGB real / 2

Table 2.8 presents the CNN architectures employed for fruit defect classification. Figure 2.8 illustrates their relative usage, showing equal prevalence of VGG16 and ResNet50, followed by VGG19, MobileNet and InceptionV3. Although reported accuracies reach one hundred percent, these results must be validated against a wider variety of defect types to ensure robustness.

Recent literature illustrates the effectiveness of convolutional neural networks for automated classification of fruits such as mangoes, apples and bananas according to quality attributes, maturity levels and geometric features. These methods frequently outperform traditional approaches by delivering high accuracy in assessing maturity and size. Nonetheless, common challenges include limited dataset sizes, inadequate benchmarking against cutting edge models and inconsistent adherence to international grading standards. Future research should focus on expanding and diversifying datasets, incorporating complementary

Table 2.8 Comparative summary of CNN models for fruit defect classification

Articles	CNN Layers (Proposed)	VGG Models	ResNet Models	MobileNet Models	Other CNN Models
[25]	–	VGG-16, VGG-19	ResNet-18, ResNet-34, ResNet-50, ResNet-101	MobileNet-V1	Xception, EfficientNet
[42]	–	VGG-19	ResNet-50	–	Xception, InceptionV3
[147]	–	VGG-16	–	MobileNet-V2	InceptionV3, DenseNet121
[162]	–	–	ResNet-50	MobileNet-V1	DenseNet121, ShuffleNet
[1]	3	VGG-16, VGG-19	–	–	AlexNet
[48]	–	VGG-16	–	MobileNet-V1	InceptionV3
[77]	2	–	–	–	AlexNet
[133]	3	–	–	–	–
[139]	6	–	–	–	–
[195]	–	VGG-16	ResNet-50	MobileNet-V3	–
[6]	–	VGG-19	ResNet-50	–	DenseNet121
[8]	–	–	–	MobileNet-V1	–
[22]	–	VGG-16	–	–	–
[101]	14	–	ResNet-50	MobileNet-V1, MobileNet-V2	InceptionV3
[111]	6	–	–	–	InceptionV3, GoogLeNet
[205]	2+LSTM	–	–	–	–
[51]	9	–	–	–	–
[81]	15, 16, 18	–	–	–	–
[132]	9	VGG-16	ResNet-50	–	AlexNet, InceptionV3
[148]	9	–	–	–	–
[150]	3	VGG-16, VGG-19	–	MobileNet-V1	Xception
[215]	11	–	–	–	AlexNet
[90]	–	–	ResNet-50	–	–

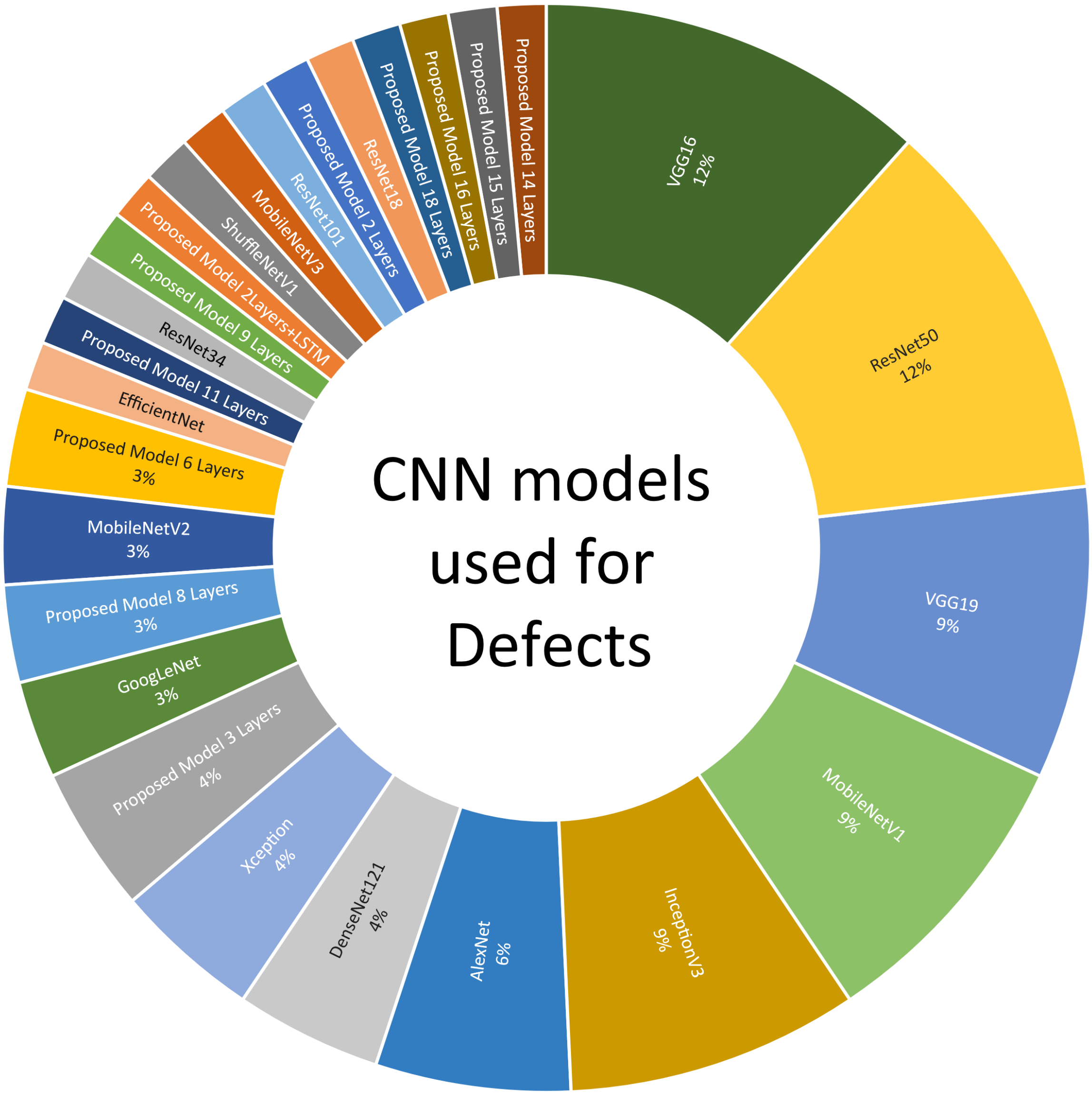


Fig. 2.8 The percentage distribution of CNN model usage for identifying fruit defects.

machine learning techniques and adopting comprehensive evaluation metrics to improve model robustness and generalizability.

2.5 Common Metrics

In summary, state-of-the-art studies on external fruit quality inspection universally adopt evaluation metrics such as Accuracy, Precision, Recall and F1 Score (see Table 2.9). To enable clear and consistent comparisons across ripeness, deformation and defect detection tasks, this proposal prioritizes Accuracy as the principal performance measure, taking advantage of its widespread acceptance and straightforward interpretation under balanced class distributions. Complementary metrics, including Precision to quantify the proportion of correct positive predictions, Recall to assess the true positive detection rate and the F1 Score to balance these two dimensions, are also reported to provide a comprehensive view of model effectiveness.

Table 2.9 Common evaluation metrics for intelligent models

Parameters	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Ripeness	✓	✓	✓	✓
Deformities	✓	✓	✓	✓
Defects	✓	✓	✓	✓

2.6 Conclusions

In conclusion, this review highlights the growing adoption of convolutional neural networks for post-harvest external quality inspection of fruits while identifying critical gaps in alignment with international grading standards and the scarcity of robust datasets with at least one thousand images per class. By analyzing literature from 2015 onward, the review synthesizes key insights into network architectures, dataset characteristics, and evaluation metrics. Leading models such as ResNet50 and VGG16 demonstrate strong performance in defect detection, and VGG16, ResNet50, VGG19 and InceptionV3 excel in ripeness classification with accuracies between 81.75% and 99.44%. Deformation classification also achieves high accuracy up to 99.04% but remains comparatively underexplored. The limited use of multispectral imaging data and the potential of synthetic data generation emerge as promising directions for expanding dataset diversity and improving model generalization. These findings establish a comprehensive foundation for future research focused on standardizing evaluation protocols, enhancing dataset quality, and advancing DL-based fruit quality grading.

Chapter 3

Ripeness Quality

The evaluation of ripeness quality in bananas has been addressed through the development of deep learning approaches designed to operate under both controlled and challenging visual conditions. In the first study, a simple CNN model was trained with a combination of real and synthetic datasets to classify banana ripeness levels, demonstrating that data augmentation can effectively mitigate the scarcity of annotated samples and improve the accuracy of ripeness grading in automated inspection systems [35]. Building on this foundation, a subsequent study systematically assessed the robustness of various deep learning architectures under varying illumination conditions, revealing that although models achieve high accuracy under ideal lighting, their performance declines significantly when exposed to lighting variability. The integration of gamma-based augmentation was shown to partially address these limitations, enhancing feature invariance and generalization across diverse illumination scenarios. Together, these contributions establish a methodological framework that combines dataset enrichment and robustness analysis, advancing the reliability of automated banana ripeness classification for real-world applications [37].

3.1 Introduction

This section is grounded in the author's published journal article *Assessing Deep Learning Model Robustness for Banana Ripeness Classification under Varying Illumination Conditions*, which establishes the experimental framework for dataset generation, illumination enhancement and model evaluation. The procedures described here strictly follow the methodology presented in that study, ensuring consistency and rigor in assessing DL performance across a range of lighting scenarios.

Ripeness assessment in bananas underpins quality control and market value, influencing harvest timing and post-harvest handling methods [88, 93, 132]. Harvesting at the physiologically green stage extends shelf life during transport [75], while international shipping requirements impose strict ripeness thresholds for long-distance distribution [5, 197]. Distribution strategies then employ ethylene regulation or color reference scales to adapt fruit maturity to market needs [145], and ripeness ultimately guides consumer choice between cooking and fresh consumption preferences [54]. Manual inspections remain slow and prone to error [166], motivating the integration of machine vision and DL models for rapid, objective ripeness classification and digitalized decision making across the supply chain [33, 38, 99].

External quality inspection of bananas traditionally relies on visual assessment of peel colour, spotting and surface condition by trained personnel following USDA or OECD guidelines [175]. Manual methods suffer from subjectivity and inefficiency, motivating the development of automated DL techniques [87].

DL models such as convolutional neural networks and vision transformers (ViT) have demonstrated strong performance in post-harvest ripeness classification. Saranya et al. [177] train a lightweight custom CNN on 273 banana images spanning four maturity stages and achieve 96.14% validation accuracy, but limited data volume and class imbalance restrict its generality.

Chuquimarca et al. [35] construct a hybrid real and synthetic banana dataset and apply transfer learning to refine a simple CNN, outperforming ResNet50, VGG19, InceptionV3 and InceptionResNetV2 with 91.70% accuracy; however, image count imbalances and uncontrolled environmental factors remain unaddressed.

Knott et al. [97] leverage a pretrained ViT to match the best CNN accuracy within one percentage point using up to three times fewer samples, highlighting Green AI benefits, though transformer tuning and adaptation to low resource settings require further study.

Arunima et al. [7] capture 4,320 Nendran banana images in a controlled chamber to train a nine layer CNN that outperforms VGG16, VGG19, InceptionV3, ResNet50 and EfficientNetB0 with 95% accuracy, yet the model's resilience to uneven illumination is not evaluated.

Chang et al. [30] introduce a hybrid physics-based and learning-based color constancy algorithm to preprocess palm fruit images before YOLOv8 detection, boosting mean average precision by 1.5 percentage points under variable lighting and underscoring the value of illumination normalization.

Nuanmeesri [141] propose a hybrid attention convolutional network for avocado ripeness classification on resource constrained devices, combining spatial channel and self attention modules via transfer learning on EfficientNetB3 and MobileNetV3 to exceed 91% accuracy while preserving inference speed and memory efficiency.

Correcting low light images is essential because ripeness estimation depends on the green to yellow colour ratio and browning patterns. Classical histogram methods and recent DL approaches often oversaturate channels and distort colourimetry [59, 84, 73]. Table 3.1 compares leading illumination enhancement techniques with respect to training requirements limitations and preservation of green to yellow balance critical for banana ripeness estimation.

Table 3.1 Comparison of recent low-light enhancement methods and their suitability for banana ripeness classification

Method	**Training required**	**Main limitations**	**Key advantage for this study**
LIME [60]	No	Slight edge sharpening and no curve fine-tuning	Preserves the green–yellow chromaticity critical for ripeness assessment
Retinex-Net [212]	Yes	Requires paired low/high-light images and a large labelled dataset	Recovers fine detail in shaded regions
Zero-DCE [59]	Yes	Sensitive to uniform textures; needs thousands of images to generalise	Compact network with pair-free training
EnlightenGAN [84]	Yes	Over-saturation in R and G channels; large unpaired dataset needed	Global perceptual enhancement while preserving structure
LLFlow [214]	Yes	High VRAM usage; extensive data required for fine-tuning	State-of-the-art visual quality with consistent brightness via invertible flow
SCI [124]	Yes	Manual curve adjustment per batch; requires wide lighting diversity	Fast GPU inference with pair-free training

DL based low-light enhancement methods such as Retinex-Net, Zero-DCE, EnlightenGAN, LLFlow and SCI achieve high perceptual quality but depend on large annotated datasets and can introduce color distortions like oversaturation or texture smoothing. Low-light Image Enhancement (LIME) requires no training data, preserves the green to yellow color balance needed for accurate ripeness assessment and features a lightweight implementation, making it the most practical and cost effective choice for enhancing banana images captured under uncontrolled lighting.

De-Arteaga et al. [41] highlight two principal challenges in post-harvest vision based quality assessment. Dataset limitations arise from the absence of standardized acquisition protocols along the production chain which leads to lighting variations, poor sample representativeness and unpredictable environmental variability. Resource constraints demand that models execute efficiently on common devices such as mobile phones or low cost cameras since farmers and packing operators often cannot afford specialized hardware. This thesis addresses these challenges by employing lightweight neural architectures optimized for real time performance on commodity hardware and by extending the controlled illumination augmentation strategy of Chuquimarca et al. [35] to rigorously evaluate model robustness under realistic lighting changes.

3.2 Banana Ripeness

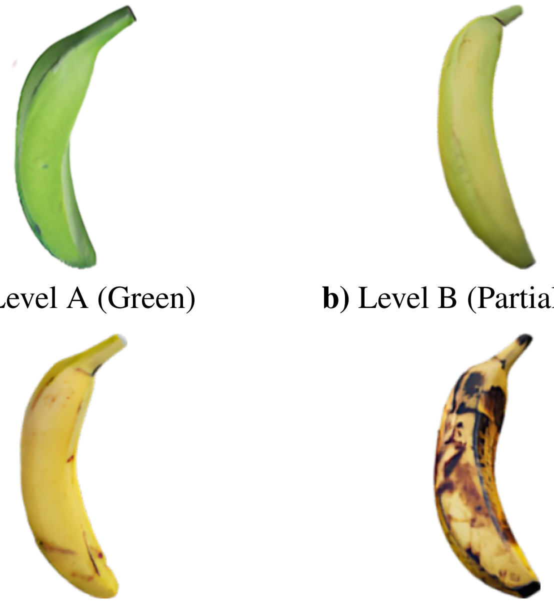
Banana ripeness is a primary determinant of fruit quality, shaping post-harvest handling, pricing and consumer acceptance. Accurate, standards aligned classification ensures consistent grading and marketability, matches maturity to logistics and end use, and extends shelf life through appropriate storage and distribution practices [12].

Conventionally, bananas are graded by peel coloration and surface appearance. Although seven stage schemes exist, industry practice often consolidates them for clarity. In this study, four levels aligned with market requirements are adopted [35] (see Fig. 3.1): Level A (Green), with a completely green peel, firm texture and predominantly starch composition; Level B (Partially ripe), with a mostly green peel showing initial yellow areas, reduced firmness and the onset of starch to sugar conversion; Level C (Ripe), with a fully yellow peel, optimal softness for immediate consumption and predominantly sugar composition; and Level D (Overripe), with extensive brown spotting, soft texture and high sugar content, suitable for immediate consumption or processing.

3.3 Dataset Generation

This study uses the public dataset at https://github.com/luischuquim/BananaRipeness, which contains images of bananas across four ripeness stages and includes both real and synthetic samples [35]. While synthetic images broaden the training space and support model pretraining, our analysis prioritizes real images to capture natural variability in peel appearance under diverse lighting and maturation conditions, thereby providing a more faithful basis for training and evaluation of DL models.

The dataset comprises 3,495 real images of Cavendish bananas acquired in a controlled climate chamber at temperatures between 15 °C and 18 °C over 28 days (see Table 3.2), approximating the full post-harvest ripening trajectory [167]. Four ripeness levels are defined at weekly intervals. The acquisition protocol produced an imbalanced distribution across



a) Level A (Green) **b)** Level B (Partially ripe)

c) Level C (Ripe) **d)** Level D (Overripe)

Fig. 3.1 Banana image samples representing the four ripeness levels used in this work: a) Level A (Green), b) Level B (Partially ripe), c) Level C (Ripe), and d) Level D (Overripe).

stages due to practical constraints of longitudinal sampling, which can reduce training efficacy and complicate fair evaluation of DL models.

Table 3.2 Banana maturity levels per day

Duration	**Level of maturity**
1–6 days	A
7–14 days	B
15–22 days	C
23–28 days	D

Constructing a real banana image corpus is resource intensive, requiring continuous oversight across full ripening cycles and tight control of temperature and illumination. Achieving sample sizes suitable for deep learning necessitates repeated acquisition rounds, inflating operational costs and staff time. To mitigate these constraints, this study augments real imagery with photorealistic synthetic samples generated via rendering pipelines (for example, Unreal Engine), expanding coverage at each ripeness level, systematically controlling scene variability, improving class balance, and accelerating training and evaluation [80, 225].

Synthetic datasets offer a cost effective way to expand training corpora with high quality, photorealistic images that closely approximate real data [17, 147]. In this work, banana images at specific ripeness stages are rendered in Unreal Engine by conditioning a calibrated 3D model on the target level and texturing it with layers derived from real imagery to narrow the domain gap [35]. The virtual scene comprises three parallel rails (rail-1, rail-2, rail-3), each equipped with cameras at 30 distinct poses (C1–C30), enabling systematic control of viewpoint, illumination and background while producing 2D RGB frames per stage (see Fig. 3.2). All synthetic images are exported at 224×244 pixels to match the real dataset, facilitating seamless mixing, improved class balance and more robust training and evaluation.

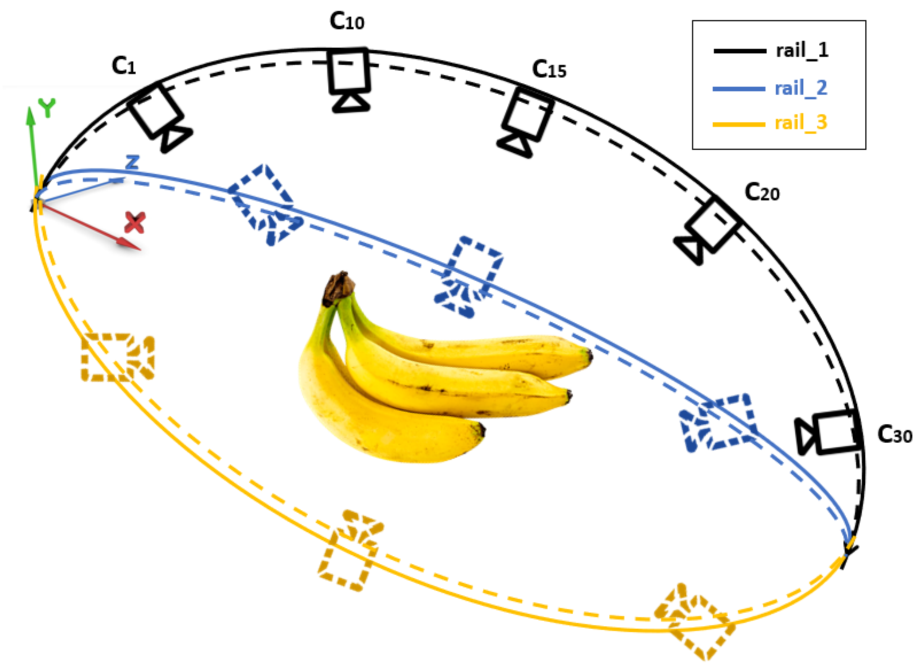


Fig. 3.2 Virtual scenario for the generation of synthetic images using Unreal Engine.

To increase domain variability, the virtual scene cycles through eight backgrounds (orange, purple, brown, light blue, asset platform, basic wall, concrete tiles, and rock marble). Across all combinations the pipeline yields 161,280 synthetic images, approximately $40\times$ the real dataset, with 40,320 images per ripeness level (see Table 3.3). For the evaluation of the models in all cases, a dataset distribution of 80% train, 10% test, and 10% validation is used.

Illumination variability is a major confounder in RGB based ripeness assessment because lighting shifts distort the green to yellow chromatic cues used for postharvest decision making. Replicating the full range of real lighting conditions is costly and time consuming, which limits dataset representativeness. To mitigate this, the real dataset is augmented with controlled low light enhancements using LIME, generating multiple brightness variants per image while preserving colorimetry. Applying this intervention only to real images retains complex surface properties and maturation patterns that synthetic renders may not capture,

Table 3.3 Summary of the banana dataset and CIDIS model performance.

Banana Ripeness Level	Number of Real Images	Number of Synthetic Images
Level A	1,429	40,320
Level B	815	40,320
Level C	559	40,320
Level D	692	40,320

broadens the training distribution without additional data collection, and improves model generalization under variable illumination in operational settings.

3.4 Lighting Variation Model

Images captured in low light exhibit reduced signal to noise ratio, low contrast and poor color constancy, which degrades perceived quality and weakens the discriminative cues available to deep learning models [108, 213]. To counter these effects, this study employs the LIME method, which estimates a pixel wise illumination map and applies structure aware enhancement to increase brightness while preserving the chromatic relationships that are critical for ripeness classification [60].

LIME augments the real banana dataset with controlled illumination variants, increasing both the volume and diversity of training samples for ripeness classification. The method estimates a per pixel illumination map by taking the channel wise maximum over R, G, and B values [60], then refines this map with structure aware regularization to enforce spatial consistency and edge preservation [219]. The enhanced illumination map is subsequently applied to the original images to improve visibility while maintaining stable colour relationships relevant to maturity assessment.

Applying LIME to the real dataset increases sample diversity through controlled illumination variants, which strengthens the ability of deep learning models to generalize across heterogeneous lighting conditions [122]. Prior studies report that LIME attains superior image quality and faster processing than other state of the art enhancers, leading to measurable gains in detection, recognition and classification tasks [94, 213]. Consequently, its use contributes directly to the robustness and versatility of the models developed for classification under real and variable conditions.

LIME first computes an initial per pixel illumination map as the channel wise maximum over the R, G, and B channels. This map is then refined with a structure aware smoothing model that preserves edges and spatial coherence and is solved efficiently in the Fourier domain using Fast Fourier Transforms. The corrected image is obtained by normalizing each RGB channel with the refined illumination map and applying a gamma correction to control output brightness. By varying the gamma parameter, multiple illumination variants of the same scene are produced, which enhances visibility and expands the volume and diversity of training data for deep learning models.

Algorithm 1 outlines the LIME based augmentation pipeline. For each image I_n in the corpus, the procedure generates multiple illumination variants by sweeping a predefined set of gamma values $\Gamma = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. Larger γ produces brighter renderings, whereas smaller γ yields subtler enhancements. This strategy expands intra class variability in a controlled manner and provides a richer basis for training and evaluating ripeness classifiers under diverse lighting conditions.

Algorithm 1 LIME applied to banana dataset augmentation

Require: Original dataset images $\{I_n\}_{n=1}^{N}$, gamma values $\Gamma = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$
Ensure: Augmented dataset images with illumination variations
for $n = 1$ to N **do**
    $I \leftarrow$ Load image I_n
    for γ in Γ **do**
        Initialize illumination map $\hat{T}$ as the maximum of R, G, B channels per pixel
        Refine $\hat{T}$ by applying structural constraints:
        Solve the following iterative optimization:
        $T^{(t+1)} \leftarrow$ Optimize illumination map
        $G^{(t+1)} \leftarrow$ Update gradient subproblem
        $Z^{(t+1)} \leftarrow$ Update dual variable
        $u^{(t+1)} \leftarrow \rho \cdot u^{(t)}$
        Repeat until convergence is reached
        Apply gamma correction: $T \leftarrow (T^{(final)})^{\gamma}$
        Enhance image: $I_{\text{enhanced}} \leftarrow I/T$
        Normalize and convert I_{enhanced} to `uint8` format
        Save enhanced image with identifier according to γ
    end for
end for

For each γ, the algorithm estimates an initial illumination map $\hat{T}$ as the channel-wise maximum per pixel on R, G, and B, which serves as the baseline for refinement. The map is then improved by solving a structure preserving optimization that enforces smoothness in homogeneous regions while retaining edges. The solver proceeds iteratively, updating the illumination estimate $T^{(t+1)}$, the auxiliary gradient variable $G^{(t+1)}$, the dual variable $Z^{(t+1)}$, and the penalty parameter $u^{(t+1)}$ until convergence, typically declared when the Frobenius norm of successive updates falls below a fixed tolerance.

After refinement, a gamma correction is applied to the illumination map to control output brightness, using $T \leftarrow \left(T^{\text{final}}\right)^{\gamma}$. The original image is then enhanced by pixel wise normalisation with the gamma corrected map, $I_{\text{enhanced}} = I \oslash T$, which markedly improves visibility in low light while preserving chromatic relationships relevant to ripeness assessment. The resulting images are subsequently normalised, converted to 8 bit unsigned integer format, and stored with filenames that encode the applied γ value for reproducibility.

This procedure yields an augmented corpus with precisely controlled illumination variants, broadening intra class diversity and strengthening the robustness and generalization of deep learning models for banana ripeness classification across heterogeneous lighting conditions. The distribution of variants and representative examples are reported in Table 3.4 and Fig. 3.3, respectively.

Table 3.4 Number of real banana images per gamma level (Γ), with brightness progressively enhanced using the LIME model.

Banana Ripeness Level	$\Gamma = 0.1$	$\Gamma = 0.2$	$\Gamma = 0.3$	$\Gamma = 0.4$	$\Gamma = 0.5$	$\Gamma = 0.6$	$\Gamma = 0.7$	**Total Images**
Level A	1,429	1,429	1,429	1,429	1,429	1,429	1,429	10,003
Level B	815	815	815	815	815	815	815	5,705
Level C	559	559	559	559	559	559	559	3,913
Level D	692	692	692	692	692	692	692	4,844

3.5 Transfer Learning

Transfer learning is employed to reduce reliance on large real datasets by exploiting synthetic imagery. Models are first pretrained on extensive synthetic corpora to learn stable color, texture, and morphology representations, then fine tuned with real images to adapt to sensor characteristics, illumination, and surface variability, and finally validated on held out real data. This two stage regimen reuses weights acquired during pretraining [218] and improves generalization and accuracy in classifying banana ripeness levels, as illustrated in Fig. 3.4.

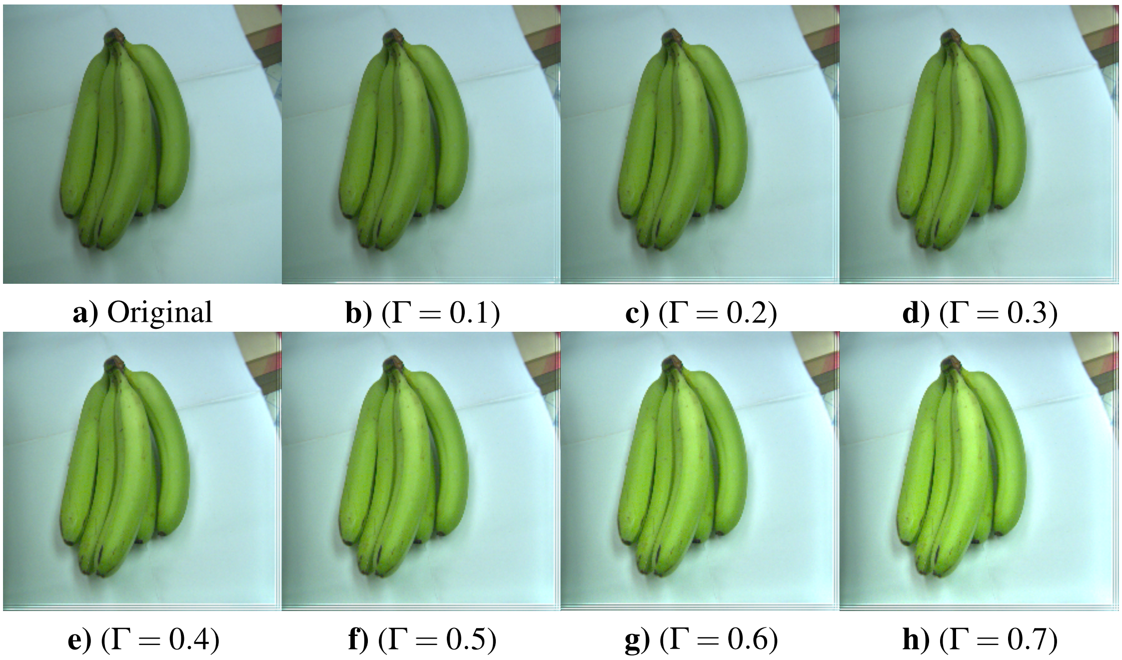


Fig. 3.3 Banana images showing progressive brightness enhancement as the gamma value (Γ) increases from 0.1 to 0.7, using the LIME model.

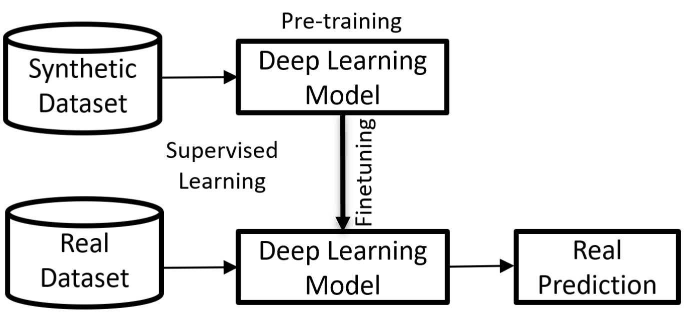


Fig. 3.4 Transfer learning using synthetic dataset.

3.6 Deep Learning Models

In this study, deep learning models are applied to extract discriminative features for external quality inspection of bananas, with the objective of classifying four maturity levels. The approach is grounded in prior scholarly review and best practices for image based fruit assessment [36].

Models are selected through a survey of the state of the art and trained on level specific image sets curated in accordance with international inspection standards. Transfer learning and controlled dataset variability are employed to improve generalization and accuracy across the four ripeness classes.

3.6.1 CIDIS Model

The CIDIS model has been progressively developed by the research group of the CIDIS Research Center at ESPOL, with the aim of proposing an efficient and low-complexity architecture oriented toward visual classification tasks. Throughout experimental evaluations, this model has demonstrated high competitiveness, achieving solid and consistent results despite its simple architecture. Its design prioritizes a balance between performance and computational efficiency, making it a suitable alternative for classification problems involving a limited number of categories and for scenarios with constrained computational resources.

In this study, the CIDIS architecture is composed of three sequential blocks, each formed by two convolution layers followed by a max pooling layer, with ReLU activations in all hidden and fully connected layers (see Figure 3.5). The network ingests RGB inputs of size 224×224 pixels and produces four outputs that correspond to the target classes, for example the four stages of banana ripeness.

Transfer learning is applied by pretraining on synthetic images, removing the original fully connected head, and training a new classification head on the target task. Model robustness is improved through hyperparameter tuning that includes dropout regularization, learning rate and batch size selection, and a comparison of Nadam and Adagrad optimizers. The final system is evaluated on real images, and prior evidence shows that the transfer learning configuration outperforms a model trained only on real data without transfer learning [35].

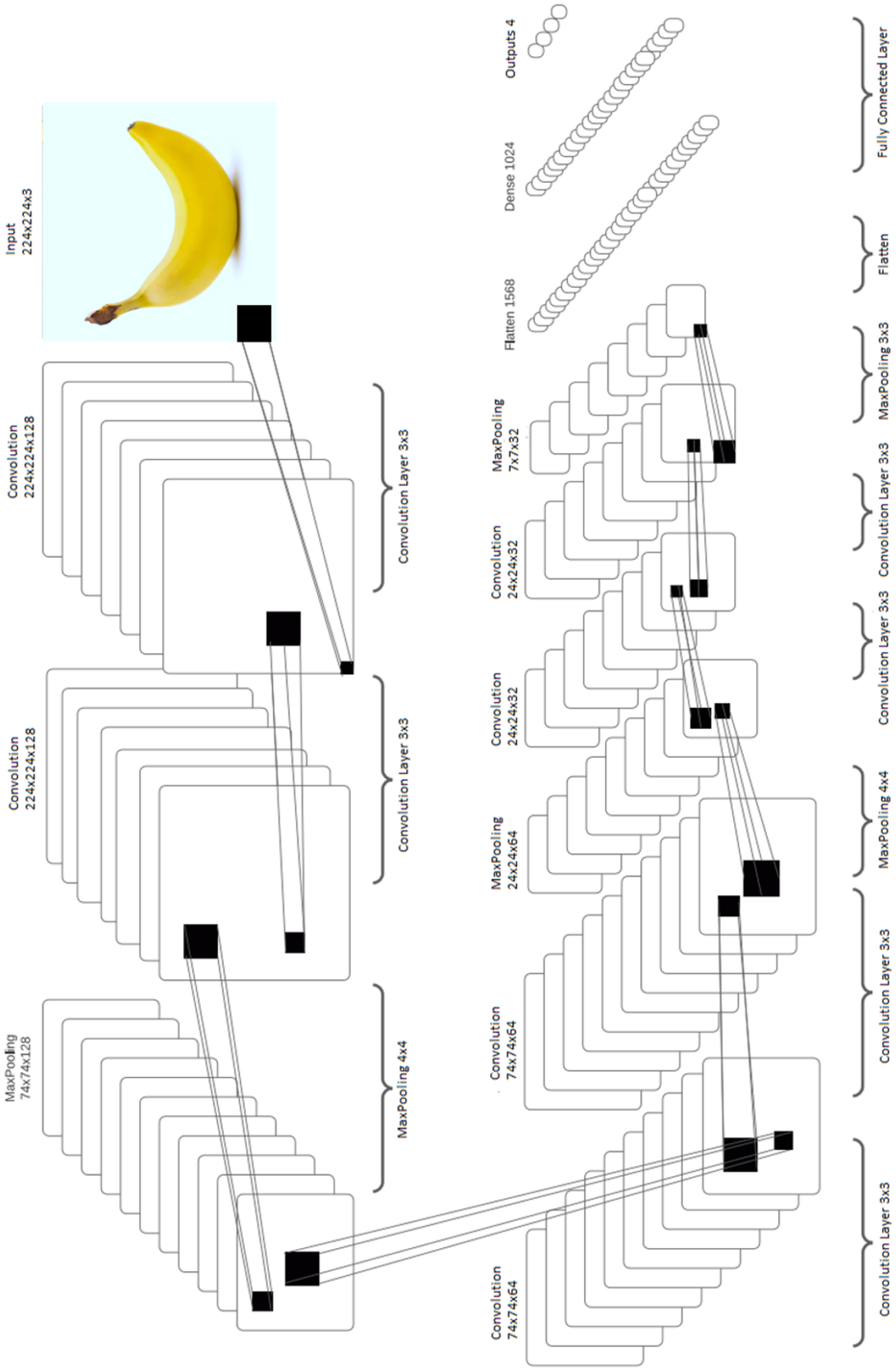


Fig. 3.5 CIDIS model architecture.

3.6.2 Vision Transformer Model

The ViT architecture has emerged as a promising alternative to conventional CNNs in image classification tasks. Recent studies demonstrate that ViTs can surpass state-of-the-art CNNs in performance, highlighting their capacity to capture global contextual information from the initial layers of the network. Unlike CNNs, which are limited by inherently local receptive fields at early stages, ViTs process images as sequences of patches, enabling a broader contextual understanding from the outset. Moreover, ViTs provide a high degree of parallelization, which enhances computational efficiency and scalability, making them particularly advantageous for large-scale image datasets [67, 92, 113].

The attention mechanism selectively emphasizes the most relevant components of the input while diminishing less informative regions. In ViT, convolutional layers are replaced by self-attention modules, enabling interactions between pixels that may be spatially distant. Through this mechanism, each element in the sequence attends to others, learning which relationships to prioritize. Unlike CNNs, which rely on inductive biases such as locality and translation invariance, ViT models operate without these constraints, allowing them to capture both fine-grained details and global dependencies across the image. This flexibility provides a richer representational capacity, although it often demands larger datasets and extended training to achieve optimal performance [20, 71].

The ViT model is composed of multiple transformer blocks, where the Multi-Head Attention (MHA) mechanism is applied to sequences of image patches. These blocks process the input into a tensor of dimensions (batch size, number of patches, projection dimension), which encapsulates the learned representations. A classification head, implemented with a SoftMax layer, subsequently transforms these representations into probability distributions across the output classes, enabling accurate prediction of the target categories.

A detailed schematic of the ViT framework used in this study is presented in Figure 3.6. Following the pioneering work of [67], which introduced the Transformer model to computer vision, the ViT architecture was developed with three principal components: patch embedding, stacked Transformer encoders for feature extraction, and a classification head. The model processes input images by dividing them into fixed-size patches, thereby reducing dimensionality and enabling efficient handling of high-resolution data. Unlike conventional CNN-based approaches, ViT employs self-attention mechanisms to capture global dependencies across patches, achieving superior representation learning and enhanced training performance even under limited data conditions.

The input image, such as that of a banana, is initially divided into fixed-size patches (e.g., 16×16), which are then embedded into encoded vectors and fed into the Transformer encoder. This encoder, composed of stacked Transformer blocks, learns feature representations from

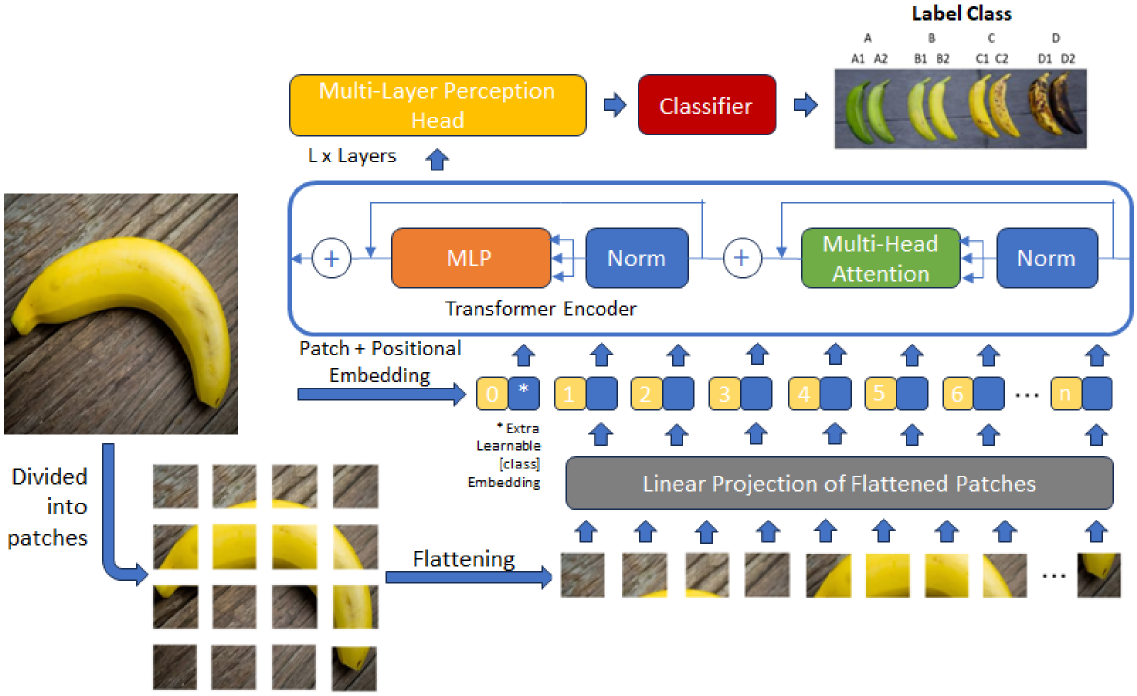


Fig. 3.6 ViT model architecture, illustrating the main components: patch embedding, stacked Transformer encoders with multi-head self-attention, and the classification head [67, 206].

the embedded patches [216]. Each block integrates MHA and a two-layer perceptron, combined with layer normalization and residual connections to ensure stable learning. The final Multi-Layer Perceptron (MLP) head produces the output, which in image classification tasks is processed by a Softmax layer to yield the predicted class probabilities. In this study, the ViT-B/16 model was employed, where B denotes the Base configuration trained on moderately sized datasets and 16 refers to the input patch size of 16×16.

The ViT model converts a sequence of image patches into semantic labels through a Transformer encoder designed for classification tasks. Unlike conventional CNNs, which rely on filters constrained by a limited receptive field, ViT applies an attention mechanism capable of focusing on different regions of an image and integrating information across its entire structure. This capability established ViT as the first image recognition model to consistently surpass CNN-based architectures, particularly those constrained by local filters. Structurally, the ViT comprises three main components: an embedding layer, stacked encoder blocks, and a classification head [221]. Its performance is further enhanced through pretraining on large datasets and fine-tuning for domain-specific applications. To accelerate training and reduce computational cost, transfer learning strategies are commonly employed in the implementation of ViT models [228].

3.6.3 VGG19 Model

The VGG19 architecture is a benchmark model in image recognition, demonstrating robust performance and high precision across diverse tasks. It is composed of 19 layers, including 16 convolutional and 3 fully connected layers, with its distinctive use of small 3×3 filters enabling deep hierarchical feature extraction. VGG19 achieves notable accuracy with relatively moderate training times when compared to more complex models such as InceptionResNetV2 [16]. Despite its effectiveness, the model presents limitations in computational efficiency due to its large number of parameters and relatively simple architecture, making it less optimal for resource-constrained applications.

3.6.4 ResNet50 Model

ResNet architectures are recognized for their capacity to train very deep networks without suffering from the vanishing gradient problem, achieved through the use of residual connections. In this study, ResNet50 with 50 layers is employed due to its ability to preserve information across layers, a limitation often observed in traditional deep networks. Residual connections facilitate the direct flow of information through the network, simplifying the training process and enhancing classification accuracy. ResNet50 is particularly well suited for tasks such as fruit ripeness classification, as it can capture high-level features from complex image data. Moreover, the integration of transfer learning with residual learning has been shown to optimize network parameters, improve generalization, and reduce overfitting, thereby strengthening the model's robustness in practical applications [70].

3.6.5 Inception-ResNetV2 Model

The Inception-ResNetV2 model integrates the strengths of Inception and ResNet architectures to achieve high efficiency and accuracy in image classification tasks. The Inception component captures information at multiple scales by applying filters of different sizes within the same layer, enabling the learning of diverse image representations. Residual connections from ResNet facilitate the training of very deep networks by preserving information flow and mitigating the vanishing gradient problem. With 164 layers, Inception-ResNetV2 ranks among the deepest models in computer vision, combining multi-scale feature extraction with residual learning. Its incorporation of 1×1 convolutions and residual connections optimizes computational resources, reduces training loss, and minimizes overfitting. These characteristics make it a robust choice for complex image recognition tasks such as banana ripeness classification, where both precision and computational efficiency are required [69, 190].

3.6.6 InceptionV3 Model

The InceptionV3 model represents an optimized evolution of the Inception family, designed to improve efficiency while minimizing computational demands without compromising accuracy. This architecture achieves superior performance compared to earlier models, such as VGGNet and InceptionV1, by employing dimensionality reduction strategies through 1×1 convolutions and by integrating multiple operations within a single layer. These design choices reduce computational complexity, accelerate training, and lower memory requirements, making the model highly resource-efficient. In the context of banana ripeness estimation, InceptionV3 has demonstrated competitive accuracy while maintaining low computational costs, positioning it as a practical alternative for agricultural quality inspection systems [104].

3.7 Performance evaluation

This section describes the evaluation strategy employed to assess the performance and robustness of deep learning models under varying illumination conditions. Standard classification metrics were applied to provide a comprehensive analysis of model behavior, particularly in scenarios affected by class imbalance [85, 64]. In addition, computational efficiency was examined through the number of trainable parameters and the average inference time per image. These indicators enable a balanced assessment that integrates predictive accuracy with practical feasibility for real-world deployment [41].

Accuracy quantifies the proportion of correctly classified instances relative to the total number of samples in the dataset, as expressed in Equation 3.1. Complementary metrics, including Precision, Recall, and F1-score, are formally defined in Equations 3.2, 3.3, and 3.4, and are widely applied in classification tasks [183]. Considering the inherent imbalance of real-world datasets and the tendency of accuracy to over-represent the majority class, which may introduce bias [68], the inclusion of these additional metrics ensures a more robust and reliable evaluation of model performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.3}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.4}$$

where TP refers to true positives, TN to true negatives, FP to false positives, and FN to false negatives.

To evaluate the computational demands of the models under varying illumination conditions, both the number of parameters and the inference time during testing were considered. These factors are critical to determine the cost-effectiveness of the models in real-world scenarios where lighting can change dynamically [149]. Model performance under such conditions depends not only on accuracy but also on the ability to process input data efficiently. Since lighting variations strongly influence classification tasks [151], robustness is reflected in a model's capacity to sustain accuracy while ensuring computational efficiency.

The number of parameters (P) determines the model's representational capacity while simultaneously influencing its computational cost during training and inference, as expressed in Equation 3.5 [192]. Larger models can capture more complex patterns but require significantly greater resources. Inference time ($T_{inference}$), defined in Equation 3.6, is a key measure of how efficiently a model processes individual inputs, which is particularly relevant for large-scale datasets and real-time applications [10]. In this study, the inference time per image was computed for each model under different lighting conditions to provide a robust assessment of their computational efficiency.

$$P = \sum_{l=1}^{L} \left(n_l^{\text{in}} \times n_l^{\text{out}} + n_l^{\text{out}} \right) \tag{3.5}$$

$$T_{\text{inference}} = \frac{t_{\text{total}}}{N} \tag{3.6}$$

where P is the total number of trainable parameters, L is the number of layers, n_l^{in} and n_l^{out} are the input and output units of layer l, $T_{\text{inference}}$ is the average inference time per image, t_{total} is the total time to process N images, and N is the number of images evaluated.

3.8 Results

This section reports the results of evaluating several deep learning models trained from scratch under varying illumination conditions for banana ripeness classification. Baseline models (M0) were initially trained on datasets without illumination variations and tested on an unaltered dataset (I0). Their performance was then assessed on seven additional test sets (I1 to I7), each incorporating incremental illumination variations defined by Γ values ranging from 0.1 to 0.7. A comparative analysis was conducted using accuracy, precision, recall, and F1-score to characterize model behavior under progressively altered lighting, providing insights into model robustness, generalization capability, and adaptability to real-world illumination variability.

Under the initial condition without illumination variation (I0), all evaluated models exhibited strong and consistent performance, attributed to the uniform lighting of the test images. This controlled environment preserved the green–yellow chromatic balance, a critical factor for accurately distinguishing banana ripeness levels. An analysis of Table 3.5 shows that under baseline conditions without illumination variation (I0), all models delivered strong and stable performance in terms of accuracy. The ViT model achieved the highest accuracy (93.13%), followed closely by ResNet50 (92.99%), while InceptionV3 obtained the lowest accuracy (89.84%). Precision, recall, and F1-score metrics under this condition were consistent with accuracy, maintaining values above 89% across most models.

Table 3.5 Performance metrics of DL models for banana-ripeness classification, trained on the baseline set M0 (no illumination variation) and evaluated on test sets I0–I7, where progressively higher gamma values ($\Gamma = 0.0–0.7$) are applied with LIME model to increase image brightness.

Illumination	Model	Accuracy	Precision	Recall	F1 Score
M0_I0		**0.9313**	**0.9310**	**0.9313**	**0.9310**
M0_I1		0.8956	0.8981	0.8956	0.8957
M0_I2		0.8383	0.8425	0.8383	0.8388
M0_I3		0.8813	0.8895	0.8813	0.8822
M0_I4	**VIT**	0.8884	0.8996	0.8884	0.8882
M0_I5		0.7210	0.7479	0.7210	0.7262
M0_I6		0.8598	0.8902	0.8598	0.8601
M0_I7		0.7396	0.8125	0.7396	0.7312
M0_I0		**0.9299**	**0.9318**	**0.9299**	**0.9304**
M0_I1		0.9227	0.9254	0.9227	0.9235
M0_I2		0.9185	0.9196	0.9185	0.9185
M0_I3		0.8970	0.9062	0.8970	0.8984
M0_I4	**ResNet50**	0.7639	0.7771	0.7639	0.7651
M0_I5		0.7983	0.8454	0.7983	0.7974
M0_I6		0.8569	0.8796	0.8569	0.8563
M0_I7		0.8169	0.8472	0.8169	0.8171

Continued on next page

Table 3.5 – continued from previous page

Illumination	Model	Accuracy	Precision	Recall	F1 Score
M0_I0	**InceptionResNetV2**	**0.9242**	**0.9246**	**0.9242**	**0.9243**
M0_I1		0.9213	0.9224	0.9213	0.9215
M0_I2		0.9027	0.9050	0.9027	0.9032
M0_I3		0.8727	0.8901	0.8727	0.8749
M0_I4		0.8870	0.9019	0.8870	0.8886
M0_I5		0.8598	0.8808	0.8598	0.8620
M0_I6		0.8469	0.8643	0.8469	0.8477
M0_I7		0.8255	0.8711	0.8255	0.8229
M0_I0	CIDIS	0.9227	0.9237	0.9227	0.9231
M0_I1		**0.9285**	**0.9285**	**0.9285**	**0.9283**
M0_I2		0.9099	0.9143	0.9099	0.9109
M0_I3		0.9013	0.9071	0.9013	0.9015
M0_I4		0.8712	0.8775	0.8712	0.8718
M0_I5		0.8755	0.8975	0.8755	0.8768
M0_I6		0.6781	0.7401	0.6781	0.6903
M0_I7		0.8255	0.8605	0.8255	0.8220
M0_I0	VGG19	0.9070	0.9062	0.9070	0.9062
M0_I1		**0.9199**	**0.9215**	**0.9199**	**0.9203**
M0_I2		0.8884	0.8937	0.8884	0.8885
M0_I3		0.7997	0.8068	0.7997	0.7998
M0_I4		0.8255	0.8628	0.8255	0.8250
M0_I5		0.8755	0.8897	0.8755	0.8744
M0_I6		0.7611	0.8224	0.7611	0.7567
M0_I7		0.6381	0.7205	0.6381	0.6498
M0_I0	InceptionV3	0.8984	0.8996	0.8984	0.8986
M0_I1		0.8727	0.8745	0.8727	0.8729
M0_I2		**0.9070**	**0.9129**	**0.9070**	**0.9082**
M0_I3		0.8870	0.8923	0.8870	0.8879
M0_I4		0.8655	0.8808	0.8655	0.8671

Continued on next page

Table 3.5 – continued from previous page

Illumination	Model	Accuracy	Precision	Recall	F1 Score
M0_I5		0.8627	0.8708	0.8627	0.8632
M0_I6		0.8512	0.8788	0.8512	0.8541
M0_I7		0.8255	0.8680	0.8255	0.8300

The ViT model achieved the highest accuracy (93.13% under I0), a result primarily attributed to its global attention mechanism, which enables the capture of broad spatial and contextual relationships across the image. This property is particularly advantageous when the training dataset is limited. By contrast, InceptionV3, which depends on localized receptive fields, obtained only 89.84% under the same conditions, reflecting a reduced capacity to generalize in this context. These findings highlight that, under ideal illumination, the preservation of color fidelity and dataset quality are decisive factors for performance in ripeness classification.

Under low illumination variations (I1 to I3, corresponding to $\Gamma = 0.1$ to 0.3), ResNet50 and CIDIS maintain robust and consistent performance, both exceeding 90% accuracy, with CIDIS reaching a peak of 92.85% in I1. In contrast, the ViT model shows a marked decline from I2 (83.83%), evidencing higher sensitivity to subtle lighting changes. VGG19 preserves acceptable performance through I2 but experiences a sharp decrease at I3, with accuracy dropping to 79.97%, indicating limited resilience to moderate illumination variations.

This behavior is largely attributable to the architectural design and generalization strategies of each model. ResNet50, through its residual connections, effectively learns stable representations that remain reliable under slight illumination shifts, thereby sustaining high accuracy. Although CIDIS does not employ residual connections, its lightweight architecture is optimized to preserve discriminative features, which explains its robustness under mild lighting variations. In contrast, the ViT model, while excelling under ideal conditions due to its global attention mechanism, is more sensitive to subtle illumination changes that disrupt chromatic consistency across patches. VGG19, lacking mechanisms to mitigate lighting variations, shows a steady decline as conditions deviate from training data, reflecting its limited adaptability to changes in the green–yellow chromatic balance essential for ripeness classification.

As illumination variation intensifies from I4 to I7 ($\Gamma = 0.4$ to 0.7), all deep learning models show marked declines across all performance metrics. This degradation arises primarily from distortions in chromatic cues, particularly the green–yellow relationship essential for ripeness classification. Excessive brightness saturates color channels, reducing discriminative information. VGG19, which relies heavily on color features, is the most affected, dropping to 63.81% accuracy at I7. ResNet50 experiences a sharp decrease at I4 (76.39%) but partially recovers at I6 and I7, likely due to residual connections that stabilize feature learning under distortions. CIDIS, despite its efficiency, lacks architectural mechanisms to handle extreme lighting, resulting in unstable outcomes, particularly at I6 (67.81%). The ViT model initially maintains robustness at I4 (88.84%) owing to its global attention mechanism but declines sharply to 73.96% at I7, reflecting its sensitivity to disrupted patch-wise coherence under uneven illumination. Inception-based models, including InceptionResNetV2 and InceptionV3, display a more gradual performance decline, retaining approximately 82% accuracy even under severe lighting shifts, suggesting that their multiscale feature extraction confers relative resilience to chromatic distortions.

Precision, recall, and F1-score metrics exhibit patterns consistent with those of accuracy across all illumination conditions, underscoring that lighting variation systematically impacts the overall performance of deep learning models. These metrics, which capture the ability to correctly classify relevant categories while minimizing false detections, are highly sensitive to distortions in the visual features guiding classification. ResNet50 maintains stable precision and recall values above 84% in most cases, except under extreme lighting conditions (I4 and I5), reflecting the role of residual connections in preserving feature consistency. In contrast, the ViT model shows greater variability, particularly in recall, which falls below 74% in I5 and I7. This reduction indicates that its patch-based self-attention mechanism struggles with local brightness inconsistencies, disrupting spatial coherence in learned representations and lowering classification accuracy. Overall, the degradation of precision, recall, and F1-score across illumination scenarios highlights the critical importance of robust feature extraction for ensuring classification reliability in non-uniform visual environments.

From a global perspective, the findings demonstrate that all evaluated models perform reliably under ideal lighting conditions (I0), maintaining high accuracy and consistent behavior across metrics. However, as illumination variability increases, particularly from $\Gamma = 0.4$ onward, most models experience a noticeable decline in performance. This trend is reflected in reduced accuracy, precision, recall, and F1-score across the majority of architectures. The varying degrees of sensitivity indicate that the generalization capacity of each model is closely tied to the diversity and distribution of lighting conditions represented during training. These results highlight the importance of incorporating controlled illumination variability into training datasets to enhance robustness and ensure effective deployment in real-world environments, where lighting is rarely consistent.

Figure 3.7 presents the accuracy performance of the ResNet50, ViT, VGG19, CIDIS, InceptionResNetV2, and InceptionV3 models under varying illumination conditions, from I0 (no variation, $\Gamma = 0$) to I7 (maximum variation, $\Gamma = 0.7$). Accuracy was selected as the primary metric for visualization, with continuous lines and distinct markers assigned to each model to enable clear comparative analysis.

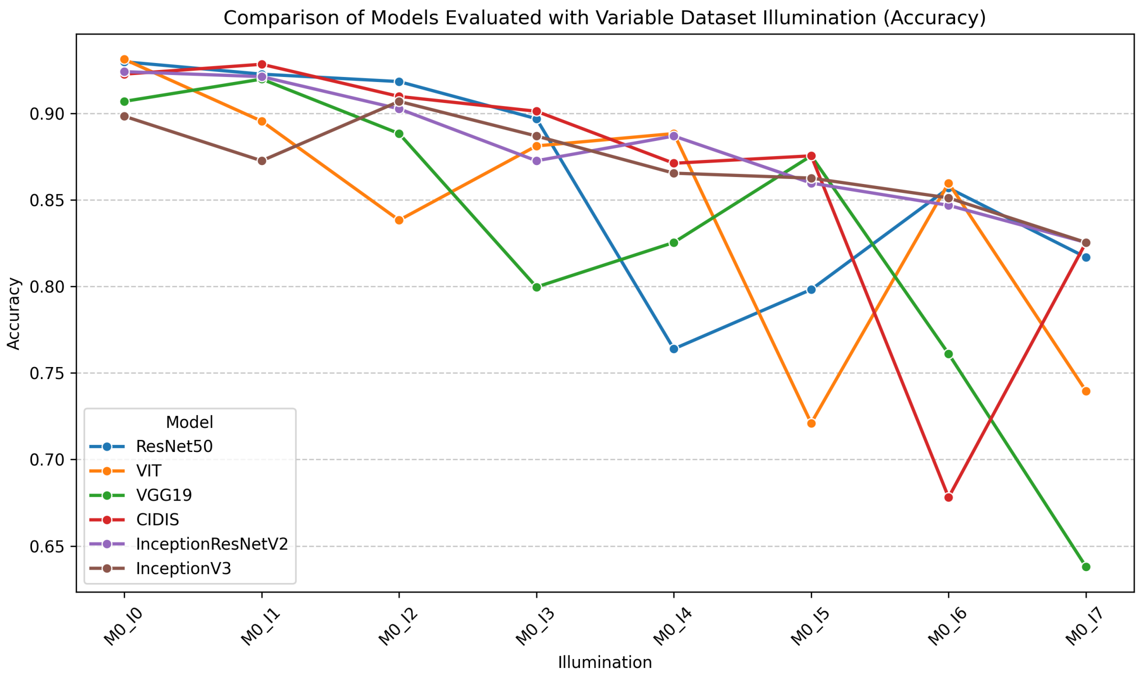


Fig. 3.7 Accuracy of DL models evaluated under illumination variation levels $\Gamma = 0.0$ to $\Gamma = 0.7$.

Under ideal conditions (I0), all models begin with high accuracy values and exhibit minimal differences. As illumination variation increases slightly (I1–I3), the performance curves start to diverge, highlighting differences in robustness. ResNet50 and CIDIS display gradual and stable declines, reflecting their resilience to moderate changes in brightness. In contrast, the ViT model shows a sharper drop in accuracy beginning at I2, indicating early sensitivity to illumination shifts. Similarly, VGG19 initially remains consistent but deteriorates rapidly at I3, underscoring its limited adaptability to even moderate lighting variability.

At intermediate to high illumination variations (I4–I7), the accuracy curves reveal distinct performance patterns across models. VGG19 exhibits the steepest and most continuous decline, reaching its lowest accuracy at I7 and confirming its limited adaptability under adverse lighting. CIDIS follows a similar trend, showing a pronounced drop from I4 to I6, but with a slight recovery at I7. The ViT model displays marked fluctuations, alternating between performance losses and partial recoveries across consecutive levels, reflecting instability under severe brightness shifts. ResNet50 experiences a sharp decline at I4 but subsequently stabilizes with modest improvements in I6 and I7. In contrast, InceptionResNetV2 and

InceptionV3 demonstrate smoother and more gradual declines, maintaining consistent behavior across all illumination levels, which highlights their relative resilience to progressive chromatic distortions.

Overall, the accuracy curves demonstrate that, although all models begin with comparable performance under ideal illumination, their trajectories diverge markedly as illumination variability increases. ResNet50 and CIDIS stand out for their stability under moderate variations ($\Gamma \leq 0.3$), reflecting stronger generalization to near-baseline conditions. InceptionResNetV2 and InceptionV3 display smoother and less abrupt declines, suggesting greater resilience under more substantial variations ($\Gamma \geq 0.4$). By contrast, the ViT model exhibits pronounced fluctuations, indicating sensitivity to both moderate and severe illumination changes. VGG19 consistently follows a downward trajectory, confirming its limited robustness across all levels of variability. This comparative analysis highlights the distinct dynamic behaviors of the evaluated architectures and provides valuable insight for selecting models according to the illumination conditions expected in practical applications.

Table 3.6 summarizes the top-1 accuracy of the evaluated models when trained on the baseline dataset without illumination variation (M0) and on the augmented dataset incorporating synthetic lighting changes (MG). Performance was assessed under two testing scenarios: I0 (no variation) and IG (gamma-augmented). This comparison enables the quantification of robustness loss when models face unseen illumination shifts and the robustness gains achieved through augmentation-based training. The outcomes are also illustrated in Figure 3.8, which provides a comparative visualization of the impact of data augmentation on classification accuracy.

The results highlight notable differences in the ability of DL architectures to cope with illumination variability. Models such as InceptionV3 show pronounced drops in accuracy, indicating greater sensitivity to conditions not represented in the training set, likely due to architectural complexity and overfitting to specific data features. In contrast, architectures like ViT and CIDIS, which integrate convolutional and dense layers, display comparatively higher resilience, although their performance also declines, suggesting limited generalization to complex input transformations. Deeper networks such as ResNet50 and InceptionResNetV2, despite their representational capacity, experience substantial reductions in accuracy, implying that the depth of their learned representations increases susceptibility to overfitting and reduces their ability to generalize to unseen lighting conditions.

The performance recovery observed when models are trained with gamma-augmented data is primarily due to the increased diversity of illumination conditions incorporated during training, which allows the architectures to learn more invariant and generalizable features. InceptionV3, initially the most sensitive to illumination variability, benefits substantially

Table 3.6 Accuracy of each DL model under illumination variation; M0 denotes models trained on non-augmented data, whereas MG refers to models trained with illumination-augmented data.

Illumination	Model	Accuracy
M0_I0	**ViT**	**0.9313**
M0_IG		0.8907
MG_I0		0.9299
MG_IG		0.9290
M0_I0	**CIDIS**	0.9227
M0_IG		0.8872
MG_I0		**0.9242**
MG_IG		0.9224
M0_I0	**ResNet50**	**0.9299**
M0_IG		0.8446
MG_I0		0.9256
MG_IG		0.9272
M0_I0	**InceptionResNetV2**	**0.9242**
M0_IG		0.8791
MG_I0		0.9156
MG_IG		0.9219
M0_I0	**InceptionV3**	0.8984
M0_IG		0.7763
MG_I0		0.9156
MG_IG		**0.9188**
M0_I0	**VGG19**	0.9070
M0_IG		0.8718
MG_I0		0.9056
MG_IG		**0.9101**

Note: The best accuracy achieved by each model is highlighted in bold, while the best overall performance across all models is additionally emphasized with a shaded background. “M0_I0” and “M0_IG” refer to models trained on non-augmented data and tested on images with no illumination change or gamma augmentation, respectively. “MG_I0” and “MG_IG” denote models trained with gamma-based illumination augmentation and tested on images with no change or with gamma augmentation.

from this augmentation, as exposure to a broader range of lighting reduces its tendency to overfit specific brightness and contrast patterns. ResNet50, CIDIS, and InceptionResNetV2 also show significant improvements, indicating that their feature extraction mechanisms adapt

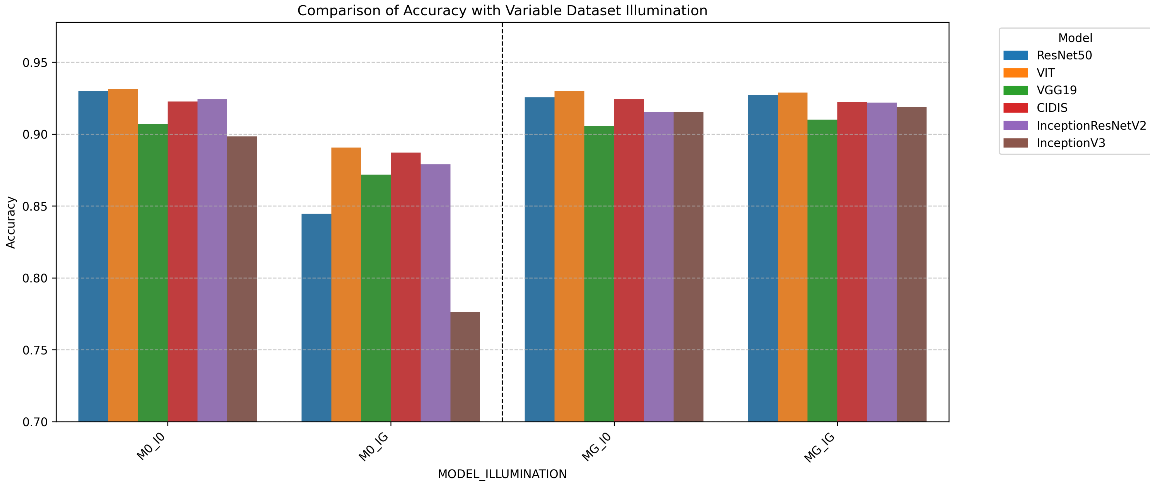


Fig. 3.8 Accuracy of the six DL models (ResNet50, ViT, VGG19, CIDIS, InceptionResNetV2, and InceptionV3) evaluated under the four illumination conditions considered in this study. M0_I0 and M0_IG correspond to models trained without illumination augmentation, whereas MG_I0 and MG_IG denote models trained with gamma-augmented data ($\Gamma = 0.0$–0.7). The dashed vertical line separates models trained on non-augmented data (left) from those trained with illumination-augmented data (right).

effectively when trained on sufficiently varied data. In contrast, the smaller gains obtained by ViT, despite its strong baseline accuracy, suggest that its attention-based mechanism already captures robust global features that are less affected by local variations in illumination.

Figure 3.8 complements the results reported in Table 3.6 by presenting the same accuracy values as grouped bar charts. This graphical representation facilitates a clearer comparison across models and training strategies, providing a more intuitive understanding of the performance differences under varying illumination conditions.

The left block (M0) highlights the susceptibility of all models when evaluated under illumination conditions absent from training. The consistent performance degradation reveals a lack of robustness, particularly in architectures prone to overfitting the original lighting distribution. While some models exhibit a comparatively milder decline, none remain unaffected by gamma-induced variation, underscoring a fundamental limitation in their generalization ability under realistic lighting variability. These findings emphasize the necessity of incorporating illumination diversity during training to reduce sensitivity to domain shifts.

The right block (MG) demonstrates the benefits of training with illumination-augmented data. All architectures exhibit substantial improvements in robustness under challenging lighting conditions while preserving strong performance under standard settings. This consistent behavior suggests that the models effectively internalize illumination-invariant features when exposed to a wider range of visual conditions during training. Moreover, the relative differences among models become less pronounced, indicating that data augmentation mitigates the vulnerabilities observed in more sensitive architectures and contributes to a more uniform performance landscape across models.

Together with the numerical evidence, the figure clearly illustrates that illumination augmentation transforms the pronounced accuracy decline observed in the M0 models into a negligible gap. This outcome emphasizes the critical role of incorporating diverse lighting conditions during training, ensuring that classifiers of banana ripeness remain robust and reliable when deployed in real-world scenarios characterized by natural variability in illumination.

Table 3.7 presents the average inference time per image, measured in milliseconds, together with the model complexity expressed as the number of parameters in millions, for each architecture evaluated on the IG test set. For every model, two configurations are reported: M0_IG, corresponding to training on the baseline dataset without illumination augmentation, and MG_IG, representing the same architecture retrained with gamma-augmented data.

Table 3.7 Average inference time per image on the IG test set and model complexity (in number of parameters, millions). Each block compares the model trained without illumination augmentation (M0_IG) versus the same model trained with gamma-augmented data (MG_IG).

Illumination	**Model**	**Avg. inference (ms)**	**# Params (M)**
M0_IG **MG_IG**	**ViT**	3.266 **2.083**	10.7
M0_IG MG_IG	VGG19	2.535 2.520	21.07
M0_IG MG_IG	CIDIS	2.689 2.664	1.9
M0_IG MG_IG	InceptionResNetV2	2.678 2.748	56.43
M0_IG MG_IG	ResNet50	3.504 3.383	26.21
M0_IG MG_IG	InceptionV3	3.891 3.790	24.42

Note: The best overall inference time across all evaluated models is highlighted in bold. M0_IG refers to models trained on non-augmented data and evaluated on gamma-augmented images (IG), whereas MG_IG refers to models trained with gamma-based illumination augmentation and evaluated on IG. “# Params (M)” indicates the total number of trainable parameters in millions.

The joint evaluation of inference time and model complexity reveals that illumination augmentation does not substantially compromise computational efficiency across the analyzed architectures. Among the models, ViT shows the most pronounced improvement, underscoring the potential of attention-based mechanisms to enhance processing efficiency when trained under variable visual conditions. ResNet50 and InceptionV3 present only modest gains, suggesting that deeper convolutional architectures benefit moderately from data augmentation without incurring additional computational costs. In contrast, CIDIS and VGG19 exhibit negligible variations in inference time, which indicates inherent robustness to illumination changes but also limited capacity for further optimization through augmentation. InceptionResNetV2 displays a slight increase in inference time, likely reflecting the higher computational demands of its hybrid architecture, where the integration of inception modules and residual connections intensifies the complexity of feature extraction under augmented training conditions.

Building on these observations, Table 3.8 presents a consolidated overview of the architectures, integrating their structural characteristics, training configurations, and performance outcomes. The Table reports the number of trainable parameters (in millions), the transfer-learning hyperparameters including optimizer, learning rate, dropout rate, and batch size, as well as the average inference times measured under non-augmented (M0_IG) and illumination-augmented (MG_IG) conditions. In addition, the column “Acc (reported)” lists the best transfer-learning accuracies on real data from [35], while “Acc M0_IG” and “Acc MG_IG” capture the test accuracies obtained without and with synthetic lighting augmentation, respectively. This integrated comparison provides a comprehensive perspective on how model complexity, computational efficiency, and predictive robustness interact, ultimately guiding the identification of the most suitable DL architecture for practical and real-time post-harvest classification tasks.

The ViT model trained with illumination augmentation (MG_IG) achieves an accuracy exceeding 92.9 % with an average inference time of 2.08 ms per image. This outcome highlights its ability to achieve an optimal balance between predictive robustness and computational efficiency, establishing it as the most effective architecture for real-time banana ripeness classification under variable lighting conditions.

LIME demonstrates superior performance compared to conventional preprocessing techniques such as CLAHE, Retinex, Color Jitter, white balance correction, and Hue, Saturation, Value (HSV) channel extraction by adaptively enhancing illumination without introducing artifacts or distorting the critical green-to-yellow chromatic balance. CLAHE often produces local over-enhancement, while Retinex and Color Jitter may alter hue distributions. White balance correction primarily addresses global tone shifts, and HSV extraction removes essential

Table 3.8 Comparative summary of model complexity, hyperparameters, inference time, transfer-learning accuracy (*Acc (reported)*) and illumination-variation performance for banana-ripeness classifiers.

Model	**#Params (M)**	**Opt.**	**Drop.**	**LR**	**Batch**	**Inf M0_IG (ms)**	**Inf MG_IG (ms)**	**Acc (rep.)**	**Acc M0_IG**	**Acc MG_IG**
ViT	**10.70**	**AdamW**	**0.1**	**1e-4**	**64**	**3.27**	**2.08**	**–**	**0.8907**	**0.9290**
ResNet50	26.21	Adagrad	0.2	1e-3	64	3.50	3.38	0.816	0.8446	0.9272
CIDIS (proposed)	1.90	Adagrad	0.2	1e-3	50	2.69	2.66	0.917	0.8872	0.9224
InceptionResNetV2	56.43	Adagrad	0.2	1e-3	64	2.68	2.75	0.869	0.8791	0.9219
InceptionV3	24.42	Adagrad	0.2	1e-3	64	3.89	3.79	0.849	0.7763	0.9188
VGG19	21.07	Adagrad	0.2	1e-3	64	2.54	2.52	0.562	0.8718	0.9101

luminance cues. In contrast, LIME operates without the need for paired training data, avoids oversaturation of sensitive channels, and preserves original color fidelity while generating realistic brightness variations in a controlled manner. These properties ensure the extraction of stable illumination-invariant features, thereby strengthening ripeness classification without compromising essential chromatic information.

3.9 Conclusion

Training classifiers on pristine images yields high baseline accuracy but exposes a critical weakness: models experience pronounced performance degradation when evaluated on gamma-altered inputs. This behavior reveals strong overfitting to clean data and limited generalization to real-world conditions, where lighting is rarely uniform. The sharp accuracy declines under variable illumination emphasize the lack of inherent robustness in these architectures, posing a significant challenge for their practical deployment in automated grading tasks.

Gamma-based augmentation markedly enhances robustness, enabling most models to recover or even exceed their baseline accuracy. Nevertheless, the improvements are uneven across architectures. InceptionV3 shows the largest relative gains but continues to underperform under severe lighting shifts. ViT, ResNet50, and CIDIS achieve consistent improvements, yet their designs remain partly ill-suited to illumination variability, as the observed gains do not fully resolve their vulnerability. These results point to structural limitations that restrict their ability to generalize under complex conditions.

Although illumination augmentation improves robustness without imposing significant computational costs, the trade-off between accuracy and efficiency remains critical. ViT and ResNet50 strike a favorable balance between predictive performance and inference time, but their relatively high computational demands could constrain real-time or resource-limited applications. While augmentation mitigates sensitivity to lighting changes, it does not completely overcome model vulnerabilities under extreme conditions, leaving open questions about their long-term reliability in dynamic environments.

The shifts in model rankings under augmented training further highlight the limitations of traditional evaluation approaches. Models such as CIDIS and InceptionResNetV2, initially weaker performers, benefit substantially from illumination augmentation, challenging the assumption that performance on clean datasets should drive model selection. These findings stress the necessity of evaluation protocols that account for real-world complexities, where illumination variability plays a decisive role in determining predictive robustness.

Chapter 4

Defect Quality

4.1 Introduction

In addressing the challenge of defect quality assessment in fruits, three complementary studies were conducted that progressively advance the state of the art in defect detection and classification. The first work, *Fruit Defect Detection Using CNN Models with Real and Virtual Data*, demonstrated the feasibility of integrating synthetic datasets with real images to enhance defect detection performance, thereby mitigating the limitations imposed by restricted training samples. Building upon this foundation, the second study, *Classifying Healthy and Defective Fruits with a Multi-Input Architecture and CNN Models*, introduced a multi-input convolutional approach that leveraged both global and local feature representations, significantly improving classification accuracy between healthy and defective specimens. Extending these findings, the third study, *Enhancing Apple's Defect Classification: Insights from Visible Spectrum and Narrow Spectral Band Imaging*, incorporated spectral imaging techniques to capture subtle chromatic variations, showing that narrow spectral bands provide complementary cues to conventional RGB data and strengthen defect classification under complex visual conditions. Together, these contributions provide a comprehensive framework that combines synthetic data generation, multi-input deep architectures, and spectral imaging strategies to improve the robustness, and precision of fruit defect quality assessment in practical post-harvest scenarios.

Each of these studies contributes incrementally to advancing the state of the art in fruit defect quality assessment. The first establishes the viability of augmenting limited datasets with synthetic images to strengthen CNN performance, the second demonstrates the advantages of multi-input architectures for integrating complementary visual information, and the third highlights the potential of spectral imaging to capture subtle chromatic cues that are often imperceptible in standard RGB images. Collectively, these works form a coherent progression that not only addresses the limitations of data scarcity, feature representation, and spectral sensitivity but also lays the foundation for more robust, scalable, and accurate post-harvest inspection systems.

4.2 Apple and Mango Defects

For this study, the analysis focuses on visually assessable defects in apples and mangoes, which represent the most common issues encountered during quality control (see Figure 4.1). The selected defects include rot, bruises, scabs, and black spots, each arising from different biological and mechanical factors. Rot is primarily caused by physical damage, insect attacks, or bacterial and fungal infections, manifesting as progressive darkening and loss of firmness in the skin [50]. Bruises, generally resulting from improper post-harvest handling and packing, alter the coloration and texture of the fruit without penetrating the skin, thus being classified as mechanical damage [136]. Scabs, particularly prevalent in apples, are fungal infections that produce cracked, dry, and darkened skin, rendering the fruit inedible [98]. Finally, black spots in mangoes, commonly associated with bacterial infections such as *Xanthomonas axonopodis pv. mangiferaeindicae*, manifest as star-shaped lesions that expand and darken over time, significantly reducing market value even at early stages of infection [39].

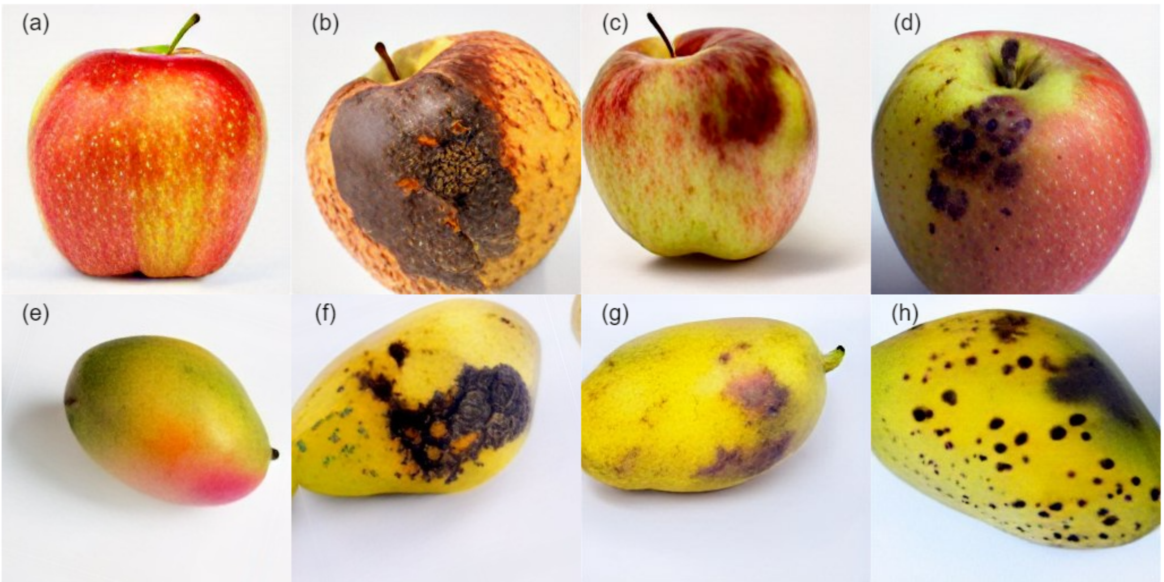


Fig. 4.1 Synthetic images of apple and mango defects generated with DALL-E mini: (a) fresh apple, (b) rotten apple, (c) bruised apple, (d) scabbed apple, (e) fresh mango, (f) rotten mango, (g) bruised mango, and (h) mango with black spots.

The defects selected for this study were determined by both the availability of training data in publicly accessible databases and the prevalence of their analysis in prior research. Rot and bruising were chosen as common defects in apples and mangoes due to their frequent occurrence and impact on post-harvest quality. Additionally, scab in apples and black spots in mangoes were included as fruit-specific defects, as they are among the most recurrent issues observed in quality control practices. This selection ensures that the evaluation covers both general and fruit-specific conditions relevant to industrial inspection tasks.

4.3 Single-Input CNN Models with Real and Synthetic Data

Food quality assurance is a fundamental requirement in the food industry, where strict standards regulate every stage of production to safeguard consumer health [142]. High-quality products must comply with acceptable sensory and physical attributes such as color, odor, texture, shape, and the absence of defects [142]. According to the FAO, between 10% and 20% of harvested fruits and vegetables in Latin American countries are discarded due to failure to meet these quality requirements [13].

The early detection of fruit defects is essential to preserve nutritional value, ensure consumer satisfaction, and minimize economic losses for producers [152]. Traditional quality inspection relies on manual evaluation, which is often slow, subjective, and prone to error, increasing the likelihood of defects being overlooked and leading to product rejection during quality control or by consumers [110, 196]. To overcome these limitations, machine learning techniques have been increasingly investigated as efficient and accurate tools for assessing fruit quality in a systematic and automated manner [45, 174].

This study evaluates the performance of CNN models for detecting defects in apples and mangoes using both real and synthetic images as training and validation data. The analysis focuses on identifying common defects such as rot, bruises, scabs, and black spots, which typically result from insect damage, plant diseases, adverse climatic conditions, or improper post-harvest handling.

4.3.1 Dataset Generation

The dataset used in this study consists of 20,000 images, equally divided between real and synthetic data. The real images were gathered from multiple online sources, including repositories, databases, and galleries, ensuring diversity in acquisition conditions. Synthetic images were generated using the DALL-E mini model, providing controlled representations of the selected defects and fresh fruit states. The dataset includes apples and mangoes with and without defects, covering rot, bruises, scabs, and black spots. Both real and synthetic datasets are publicly available for reproducibility and further research at https://github.com/luischuquim/Healthy-Defective-Fruits.

Synthetic images were generated to compensate for the limited availability of public datasets containing the selected fruit defects. The dataset was created using the DALL-E mini text-to-image model, which combines the BART language encoder with the VQGAN image decoder to generate realistic images from short textual prompts [189]. Through the HuggingFace web interface, synthetic samples were obtained for all defect categories in apples and mangoes, thereby increasing dataset diversity and supporting the training of more robust CNN models.

The heterogeneity of the collected and generated datasets required a refinement process to ensure data quality. Real images were manually inspected to remove low-quality samples and those unrelated to the target defect categories. In some cases, images were cropped to retain only the region of interest corresponding to the fruit. Due to the scarcity of real datasets for specific defects, data augmentation techniques were applied using OpenCV and Python preprocessing libraries. Random transformations such as contrast adjustments and rotations were introduced to increase variability and balance the dataset, thereby improving the robustness of the training process.

The synthetic dataset was refined through manual selection of DALL-E mini outputs to ensure relevance and visual consistency with the target defect classes. All images were standardized to 256×256 pixels, allowing efficient storage and uniformity for training. Although the experimental classification was binary, distinguishing between healthy and defective fruits, the dataset was balanced across categories to avoid bias in training. Table 4.1 presents the distribution of images per fruit and defect category, providing a comprehensive view of the dataset composition.

Table 4.1 Dataset distribution.

Fruit	**Category**	**Dataset size**			
		Real images	**Synthetic images**	**Total per category**	**Total**
Apple	Fresh	2500	2500	5000	10000
	Rot	1000	1000	2000	
	Bruise	750	750	1500	
	Scab	750	750	1500	
Mango	Fresh	2500	2500	5000	10000
	Rot	1000	1000	2000	
	Bruise	750	750	1500	
	Scab	750	750	1500	

4.3.2 Training

For model training, the dataset was organized into separate directories for apples and mangoes, each containing 10,000 images. The data were partitioned into 80% for training, 10% for validation, and 10% for testing, with the test set restricted to real images to ensure a realistic evaluation of performance.

The models selected for this study were MobileNetV2, InceptionV3, VGG16, and DenseNet121. Using the Keras and TensorFlow libraries, these architectures were loaded with pretrained ImageNet weights and subsequently adapted to the binary classification task. Specifically, the original fully connected output layers designed for multiclass classification were removed and replaced with layers tailored to the present problem.

For the binary classification task, the fully connected output layers of all models were adapted to a uniform architecture. A dense layer with 1024 units and a ReLU activation function was followed by a dropout layer, then another dense layer with 512 units and ReLU activation. The final classification was produced through a dense layer with two units and a softmax activation, yielding the one-hot encoded vector. To assess the role of regularization, three experimental configurations were evaluated: one without dropout and two with dropout rates of 0.2 and 0.5, respectively, to investigate their impact on performance metrics and mitigate overfitting.

A total of 36 training experiments were performed by combining the fruit category, CNN architecture, and selected hyperparameters. The models were trained using RMSprop as the optimizer with a learning rate of 0.001, a batch size of 16, and three different epoch settings of 10, 20, and 30. This design enabled the systematic evaluation of performance variations across architectures, datasets, and training durations.

4.3.3 Results

This section presents the evaluation results of the CNN models based on the methodology previously described. For each fruit, a total of 36 training runs were conducted, encompassing variations in CNN architectures, training epochs, and dropout configurations. This experimental design ensured a comprehensive comparison of model performance under different training conditions.

Results with the Apple Dataset

Table 4.2 summarizes the performance of the four selected models on the apple dataset, which includes images of fruits with and without defects under different dropout rates and epoch settings. The best test accuracy achieved by each model is highlighted to facilitate comparison of their classification effectiveness.

Table 4.2 Comparison of CNN model accuracy with apple image dataset.

CNN Model	**Accuracy**
InceptionV3	93.40
MobilenetV2	97.50
Densenet121	94.50
VGG16	76.20

The analysis of the training runs revealed a consistent relationship between test accuracy, the number of epochs, and the dropout rate. The highest performance was obtained with 30 epochs and a dropout rate of 0.2, which consistently improved the accuracy of the models. In contrast, omitting dropout layers or applying higher dropout rates, such as 0.5, led to reductions in accuracy, likely due to excessive information loss in the dense layers.

A clear trend emerged in which accuracy improved with longer training and a low dropout rate of 0.2. Among the evaluated architectures, MobileNetV2 achieved the highest accuracy with the apple test dataset, reaching 97.50%. InceptionV3 and DenseNet121 followed closely with 93.40% and 94.50%, respectively. Conversely, VGG16 exhibited the lowest performance, with an accuracy of 76.20%.

The evaluation of the loss metric revealed high values, indicating a substantial discrepancy between the predicted and expected outcomes. In the case of InceptionV3 and MobileNetV2, this behavior suggests the presence of overfitting, as both models achieved accuracy levels above 90% while simultaneously exhibiting elevated loss values. Overfitting is a common phenomenon in dense networks because gradient updates must propagate through the entire structure. Furthermore, the lack of additional regularization mechanisms, beyond the dropout strategy employed in this study, may contribute to this effect. As noted by [57], alternatives such as batch normalization can mitigate this limitation by enabling the network to generalize more effectively without memorizing training data.

Although these models demonstrated strong recognition performance on the training dataset, their accuracy decreased when classifying unseen images. To address this issue, potential strategies include reducing the number of training epochs and experimenting with different dropout configurations to improve the balance between model complexity and generalization capability.

Results with the Mango Dataset

For mango, the evaluation with the selected models produced results consistent with those obtained from the apple dataset. The training runs were executed using the same hyper-parameters described previously. Table 4.3 presents the performance of each model on this dataset, with the rows corresponding to the best test accuracy for each case highlighted.

Table 4.3 Comparison of CNN model accuracy with mango image dataset.

CNN Model	Accuracy
InceptionV3	91.90
MobilenetV2	92.90
Densenet121	92.50
VGG16	63.60

The best performance for mango classification was obtained with the MobileNetV2 model, reaching 92.90% test accuracy using a configuration of 30 epochs and a dropout rate of 0.2. InceptionV3 and DenseNet121 achieved their highest values with 20 and 30 epochs, respectively, both without dropout layers, resulting in test accuracies of 91.90% and 92.50%. In contrast, the VGG16 model underperformed, reaching only 63.60% accuracy in testing.

As the number of epochs increased, accuracy remained stable without abrupt variations, except in the case of VGG16. Given the complexity of this architecture, which involves hundreds of millions of parameters, greater stability may be achieved with additional training epochs. For the mango dataset, all models reached accuracies above 90% from the early epochs, reflecting a high recognition capability and a low incidence of false positives and false negatives.

Compared to the apple dataset, the loss values were consistently lower. After 30 epochs, all models reduced the loss metric to values below 0.50, indicating a more efficient convergence.

For the InceptionV3 model, however, signs of overfitting were observed, as loss values above 0.50 persisted at 20 and 30 epochs. In this case, improved results could be achieved by reducing the number of epochs or introducing dropout variations to enhance regularization.

4.3.4 Conclusions

During the dataset creation stage, a total of 20,000 images were collected, evenly distributed between apples and mangoes. This dataset included both real and synthetic images. Real images were obtained from public databases, while synthetic images were generated using the DALL-E mini text-to-image system. The dataset incorporated healthy fruits as well as defective ones, specifically representing rot, bruises, scabs, and black spots.

Since the evaluation of CNN models for fruit defect detection has been previously studied, this work proposed an enhanced dataset that improves not only in volume but also in quality. The use of DALL-E mini enabled the generation of images containing fruits with clearly defined defects, free from background elements and noise. Each defect type was represented with a balanced number of images, which allowed the CNN models to effectively learn to classify apples and mangoes according to their condition.

This dataset was then employed to evaluate four CNN architectures: MobileNetV2, InceptionV3, DenseNet121, and VGG16. The task focused on binary classification to determine whether the fruits were fresh or defective. For apples and mangoes respectively, the testing accuracies achieved were 97.50% and 92.90% with MobileNetV2, 93.40% and 91.90% with InceptionV3, 94.50% and 92.50% with DenseNet121, and 76.20% and 63.60% with VGG16. These results demonstrate that MobileNetV2 achieved the highest performance, confirming its suitability for defect detection in both fruit types.

4.4 Multi-Input CNN Models with Real Data

In post-harvest processes, fruit classification is typically performed through manual inspection by specialized agents who evaluate parameters such as ripeness, deformities, and defects [119, 151]. These evaluations adhere to international standards established by regulatory bodies, since the commercial value of fruits is directly linked to their quality level, where higher quality implies greater profitability and access to premium markets [120]. Consumers, ranging from small retailers to large supermarkets, base their purchasing decisions on several factors, although visual appearance remains the dominant criterion. In the global marketplace, ensuring strict quality control of agricultural products is essential, supported by international standards that guarantee product safety and quality for human consumption [35, 43]. Within this context, the present study focuses on the parameter of fruit defects, specifically in apples and mangoes, with the aim of classifying fruits as either healthy or defective.

In recent years, the development of automated systems based on computer vision has advanced considerably, enabling the replacement of manual manufacturing operations [29, 44]. Within this context, the classification of fruits as healthy or defective has been widely addressed using CNNs, which process fruit images as input [109, 135]. These models evaluate images by applying weights derived from the training process, typically using binary and true-color formats as the primary forms of digital representation. To extract meaningful information from these images, segmentation techniques are applied to isolate relevant features [62, 132]. Once segmented, the images are assigned labels according to specific visual attributes, such as the distinction between healthy and defective fruits [25, 77]. In this study, two CNN architectures, MobileNetV2 and VGG16, were selected for analysis to evaluate their performance in the classification of fruit defects.

The main contributions of this study can be summarized in four aspects. First, CNN models performance was improved by relying exclusively on real images, eliminating the need for synthetic image generation as in previous studies [147]. Second, image pre-processing was introduced through silhouette segmentation, a flat shape descriptor that allows for a more precise identification of defects such as rot, bruises, scabs, and black spots, which in turn significantly improved improved cation accuracy. Third, a novel multi-input architecture was implemented, integrating both RGB images and silhouettes into separate branches, thus surpassing the results obtained in previous single-input approaches [147]. This architecture enabled a more effective extraction of distinctive fruit features and yielded higher classification accuracy. Finally, the proposed methodology combined RGB images with their corresponding silhouettes to train multi-input models using MobileNetV2 and VGG16. As a result, MobileNetV2 achieved remarkable accuracy, reaching 100% in classifying healthy and defective apples and 99.53% in mango classification. These results clearly outperform those reported in earlier studies, where MobileNetV2 reached 97.50% for apples and 92.90% for mangoes. This comparative analysis highlights the effectiveness of the proposed architecture and provides valuable insights for future research on fruit defect detection.

4.4.1 Data Generation

The dataset employed in this study was obtained from a public repository (https://github.com/luischuquim/Healthy-Defective-Fruits) and consists of real and synthetic images of apples and mangoes. For the experiments, only 10,000 real images were used, comprising an equal distribution between healthy and defective samples for both fruit types. The dataset is carefully annotated, ensuring consistency and reliability in the classification process.

The acquired dataset was processed using the Segment Anything Model (SAM) tool [46] to generate fruit silhouettes from the original images. This segmentation enables the extraction of shape descriptors, providing valuable structural information for classification. In healthy samples, the silhouettes are binarized, showing the complete fruit contour in white. In defective samples, the silhouettes are also binarized, but defects appear as dark pigmentations within the white fruit region, reflecting visible irregularities (see Figure 4.2 and Figure 4.3).

After compiling the dataset with 5,000 images per fruit, the SAM model was applied to generate segmented silhouettes. Approximately 25% of the images were segmented incorrectly, highlighting inconsistencies in the automated process. To address this issue, a

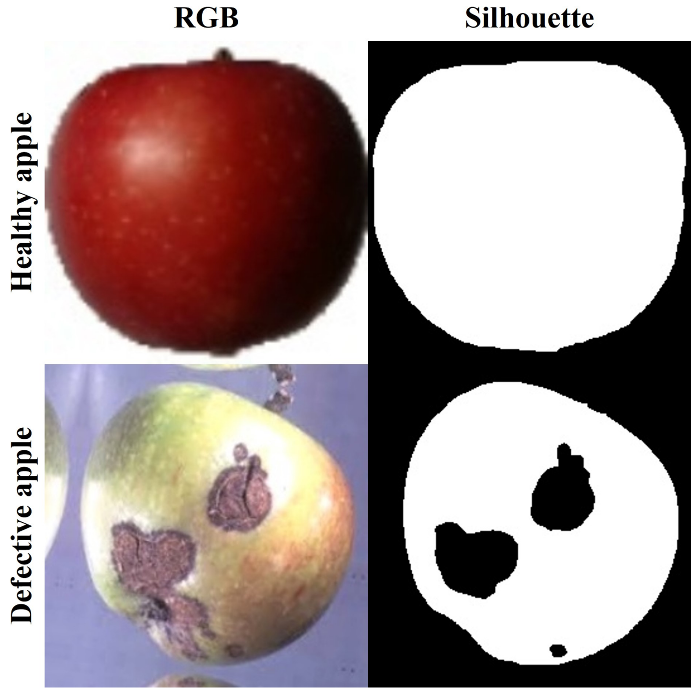


Fig. 4.2 Binarized silhouette of an apple obtained through segmentation.

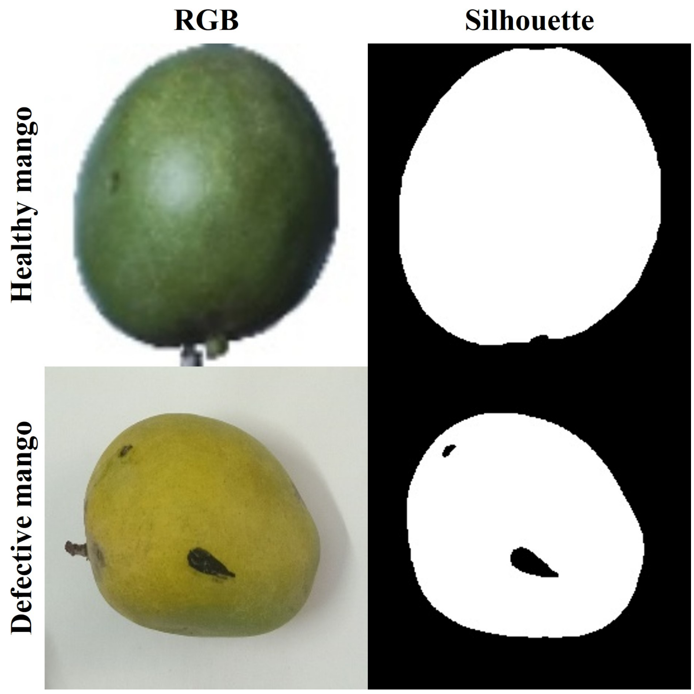


Fig. 4.3 Binarized silhouette of a mango obtained through segmentation.

manual refinement was conducted, ensuring that only accurately segmented images were retained. The final distribution of images used in this study is summarized in Table 4.4.

Table 4.4 Dataset distribution.

Fruits	Category	Evaluation	# Original Images	# Refined Images
Apple	Healthy	Train	2000	1823
		Validation	250	218
		Test	250	214
	Defective	Train	2000	1459
		Validation	250	170
		Test	250	173
Mango	Healthy	Train	2000	1624
		Validation	250	203
		Test	250	221
	Defective	Train	2000	1478
		Validation	250	177
		Test	250	202

4.4.2 Multi-Input architecture

The multi-input architecture integrates multiple data streams, each processed by an independent CNN model, and subsequently merges the extracted features to perform the classification task [32, 47, 126]. In this implementation, two distinct branches were considered: one processing RGB images and the other their corresponding silhouettes. Each branch employed a pretrained CNN model, specifically MobileNetV2 and VGG16, trained independently to capture complementary visual representations. The outputs from both branches were fused through a feature integration stage, followed by a MLP layer that completed the classification. The methodology adopted in this study is illustrated in Figure 4.4.

4.4.3 Training

For training, validation, and testing, the dataset was distributed following an 80%, 10%, and 10% partition, respectively, as summarized in Table 4.4. The images were initially organized into directories by fruit type, with 5,000 images each for apples and mangoes. After refinement, the final dataset comprised 4,057 apple images and 3,905 mango images, ensuring balanced and representative samples across categories.

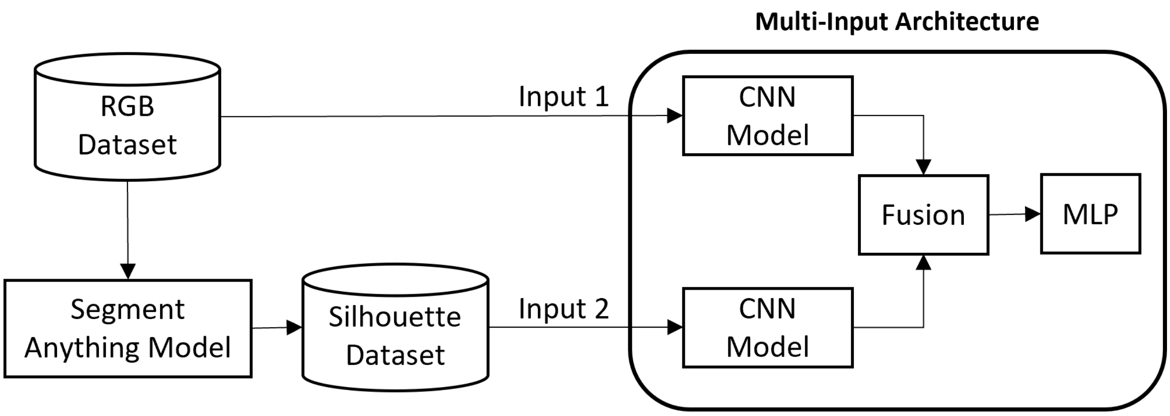


Fig. 4.4 Multi-Input architecture used for the classification of healthy and defective fruits.

Based on the state-of-the-art review, the convolutional neural network models selected for this study were MobileNetV2 and VGG16, both widely recognized for their effectiveness in image classification tasks.

MobileNetV2 is specifically designed for computer vision tasks on resource-constrained devices, such as mobile phones, which makes it suitable for applications involving fruit inspectors, producers, or consumers. This architecture achieves high efficiency by employing Inverted Residual Blocks, which reduce the number of parameters and computational operations while maintaining strong classification performance, particularly in distinguishing healthy from defective fruits. Furthermore, MobileNetV2 incorporates the Expansion Factor technique to regulate the number of channels in the intermediate layers, enhancing its adaptability to diverse tasks and computational requirements [217].

VGG16 is a deep architecture composed of 16 layers of convolution and pooling, followed by three fully connected layers, characterized by its simplicity and relative depth. The model applies small filters of size 3×3 in all convolutional layers, with max-pooling operations to progressively reduce dimensionality. VGG16 has consistently demonstrated strong performance in image classification tasks on standard datasets, which aligns with its effectiveness in this study [201].

4.4.4 Results

This section presents the results obtained from evaluating Multi-Input architectures combined with CNN models, following the methodology previously described. For each fruit type, the complete dataset, including RGB images and their corresponding silhouettes, was employed to assess the models. To ensure comparability with the study by [147], the same fruit categories and defect types such as rot, bruises, scabs, and black spots were considered. Both models initially achieved 100% accuracy in fruit classification. However, after reevaluating the CNN models with the refined datasets, consistently high performance was maintained, with MobileNetV2 being the only model to reach 100% accuracy.

The hyperparameters were configured to train the architectures using input images of size 224×224. The CNN models employed pre-trained weights from ImageNet to leverage transfer learning. Training was conducted with the Adam optimizer, a learning rate of 0.00001, and a batch size of 16. Although the task involved binary classification, categorical cross-entropy was selected as the loss function due to its stability and widespread application in multiclass settings. Training was initially set for 60 epochs; however, since satisfactory performance was consistently achieved within the first 30 epochs, the process was curtailed accordingly. The final models were determined by selecting the weights corresponding to the epoch with the best validation performance.

Results with the Apple Dataset

For the classification of healthy and defective apples, two CNN models were integrated into the Multi-Input architectures, namely MobileNetV2 and VGG16. The analysis of training and validation plots demonstrated consistently high accuracy, primarily due to the inclusion of silhouette images. The segmentation process provided by silhouettes enhanced the ability of the Multi-Input architectures to reliably distinguish between healthy and defective fruits. During the testing phase, the MobileNetV2-based Multi-Input architecture achieved perfect accuracy, correctly classifying all images of healthy and defective apples. In contrast, the VGG16-based architecture exhibited limitations in identifying healthy apples. This behavior is likely related to its design, which is optimized for natural images with rich textures and detailed features. When applied to silhouettes in binary images, relevant information is reduced, limiting its capacity to adapt to the simplicity inherent in silhouette representations.

Results with the Mango Dataset

For the classification of healthy and defective mangoes, two CNN models, MobileNetV2 and VGG16, were integrated into the Multi-Input architectures. The training and validation plots revealed consistently high accuracy, mirroring the results obtained with apples. This performance is largely attributed to the inclusion of silhouette images, which demonstrated the versatility of this approach across different fruit types. During testing, the MobileNetV2-based Multi-Input architecture correctly classified all healthy mangoes and nearly all defective mangoes, resulting in excellent performance. Similarly, the VGG16-based architecture achieved high accuracy, successfully identifying almost all images of both healthy and defective mangoes, although with slightly lower precision compared to MobileNetV2.

Upon examining the performance metrics presented in Table 4.5, the MobileNetV2 Multi-Input model demonstrated the best results for the classification of healthy and defective fruits. This architecture achieved 100% accuracy for apples and 99.53% for mangoes. When compared to the results reported in the study by [147], which attained accuracies of 97.50% for apples and 92.90% for mangoes using only the MobileNetV2 model, the superiority of the proposed Multi-Input approach becomes evident. It is important to emphasize that the same real dataset from the reviewed work was employed in this study, ensuring a consistent basis for comparison. Moreover, the MobileNetV2 Multi-Input model proved to be versatile, as it is suitable not only for silhouette images but also for other types of segmented binary images, highlighting its potential applicability across diverse classification scenarios.

Table 4.5 Performance comparison of multi-input CNN models for fruit defect detection.

Fruits	Models	Accuracy	Precision	Recall	F1 Score
Apple	**MobileNetV2 (Multi-Input)**	**100.00**	**100.00**	**100.00**	**100.00**
	VGG16 (Multi-Input)	81.65	85.45	83.41	81.54
Mango	**MobileNetV2 (Multi-Input)**	**99.53**	**99.55**	**99.50**	**99.53**
	VGG16 (Multi-Input)	97.39	97.38	97.43	97.40

The results obtained with the Multi-Input architectures, particularly the MobileNetV2 model, demonstrate exceptionally high performance, raising the question of potential overfitting. However, this concern is mitigated by the fact that training and evaluation were conducted on carefully partitioned datasets using cross-validation strategies, ensuring that the reported metrics genuinely reflect the model's generalization capability rather than memorization of training samples. The inclusion of real images in the validation folds provided a robust assessment framework, confirming that the model can reliably classify unseen data. Furthermore, the consistency of results across different fruits reinforces the stability of the Multi-Input approach, showing that its ability to integrate RGB and silhouette information is not limited to a specific dataset but rather generalizable to broader classification scenarios.

4.4.5 Conclusions

The implementation of robust models demonstrated superior performance compared to state-of-the-art approaches, achieving accurate classification of healthy and defective fruits through a Multi-Input architecture that integrates RGB and silhouette datasets. The inclusion of binarized silhouettes markedly enhanced the model's discriminative capability, enabling more precise detection of defects and overall improved performance. Among the evaluated architectures, the MobileNetV2 (Multi-Input) model achieved the best results, reaching 100.00% accuracy for apples and 99.53% for mangoes, thereby establishing its effectiveness for fruit defect classification tasks.

Additionally, when compared to the previous work in which synthetic images were incorporated to increase dataset size, the present study achieved a considerable improvement in model performance by relying exclusively on real images. This performance gain is primarily attributed to the application of the SAM for accurate fruit segmentation and the adoption of a Multi-Input architecture that jointly processes RGB images and their corresponding silhouettes.

4.5 Apple Defect Classification with Visible and Narrow-Band Imaging

In the global agricultural market, quality control has become a decisive factor for both producers and consumers [66, 165]. International standards are essential to guarantee the production of high-quality fruits and vegetables, which play a critical role in safeguarding human health [130, 173]. The acceptance of products relies on attributes such as ripeness, deformation, and the presence of defects, which are carefully evaluated by inspectors and ultimately influence consumer decisions [35, 132, 199]. The economic value of food products is strongly tied to their quality, with higher categories commanding greater market prices as defined by international trade regulations. Consequently, for producers, the implementation of reliable inspection systems capable of accurately classifying fruits into appropriate categories is crucial for ensuring market access and maximizing profitability.

The identification of fruit defects is a critical step to ensure quality, preserve nutritional value, satisfy consumer expectations, and reduce financial losses for producers [147]. At present, quality inspection in the agricultural industry largely relies on manual evaluation methods [26]. These approaches are slow, prone to errors, and often fail to detect subtle imperfections, leading to rejection by quality inspectors or consumers. To overcome these limitations, machine learning techniques in computer vision have emerged as promising alternatives, enabling external fruit quality assessment with greater speed, consistency, and accuracy.

The present study evaluates the performance of deep learning architectures, specifically CNN models, for the detection of defects in apples. The analysis is conducted using a dataset composed of visible spectrum images and narrow-band images at a wavelength of 660 nm as input for training and validation. The defects of interest include rot, bruises, scabs, and black spots, which commonly arise from insect damage, diseases, climatic factors, and post-harvest handling. According to international quality standards, the presence of any defect results in the rejection of the fruit. Consequently, the classification task is formulated as a binary problem, categorizing apples exclusively as healthy or defective.

This study provides several contributions to the field of apple defect classification. A dataset was acquired and constructed, including three types of apple defects: bruises, stains, and rot, captured under both visible spectrum and 660 nm wavelength imaging. For each defect type, silhouette images were also generated. This dataset has been made publicly available to the scientific community through GitHub (https://github.com/cidis-vision/apple-defects), ensuring accessibility and reproducibility. Furthermore, the study evaluates the performance of Single-Input and Multi-Input neural network architectures in defect classification by leveraging data from multiple spectral ranges. Finally, the proposed methodology enables the classification of three major apple defects with potential applications in quality control systems.

4.5.1 Data Generation

The dataset employed in this study was acquired through a specialized imaging system composed of two parallel cameras for simultaneous capture. The first camera, MER2-160-227U3C, was used to obtain images in the visible spectrum. The second device, a Basler ace acA1300-60gmNIR, was equipped with a BP660 bandpass filter operating within a spectral range of 640–680 nm, centered at 660 nm, with a Full Width at Half Maximum (FWHM) of 60 nm. The resolutions of the two cameras were 1280×960 pixels and 1280×1024 pixels, respectively, enabling high-quality acquisition in both spectral domains.

To guarantee comprehensive coverage of each fruit, a rotating table was employed to allow sequential acquisition from multiple angles. This setup enabled the capture of 120 images per fruit, ensuring that surface variations and potential defects were fully recorded. The synchronized operation of the two cameras facilitated simultaneous acquisition in both spectral domains, providing temporally coherent datasets suitable for accurate analysis and model training.

The dataset comprises three primary defect categories: bruises, stains, and rot. Due to the simultaneous acquisition process, an equal number of images were obtained from both cameras for each fruit sample. To generate instances of bruises, six healthy apples were subjected to controlled mild impacts, with images captured across six consecutive days to record defect progression. For stains, a set of 20 apples was used, while another set of 20 apples was employed for rot. This structured acquisition ensured realistic variability in the dataset, enabling the models to effectively learn and generalize across different manifestations of the defects.

The dataset consists of 4,539 images of bruises, 3,136 images of stains, and 3094 images of rot, totaling 10,769 images (see Table 4.6). This diverse collection, acquired under different angles and defect conditions, provides a robust foundation for training and evaluating the proposed classification models, ensuring their applicability across varied scenarios. Prior to model training, two stages of image post-processing were applied: image registration to align the captures and defect segmentation to isolate the relevant regions for analysis.

Table 4.6 Dataset distribution of apple defects.

Apple Defects	**D.1 (# visible)**	**D.2 (# 660 nm)**	**D.3 (# masks – visible)**	**D.4 (# masks – 660 nm)**
Bruises	4539	4539	4539	4539
Spots	3136	3136	3136	3136
Rot	3094	3094	3094	3094

Image Registration

Following image acquisition, registration was required due to the spatial misalignment inherent in the dual-camera configuration. To ensure accurate alignment of image pairs, a registration process was implemented. The LightGlue model [114] was employed for this task, producing reliable results and achieving the necessary spatial coherence between the two image streams.

The registration stage was designed to correct spatial misalignments generated during simultaneous image acquisition, ensuring precise pixel-to-pixel correspondence between paired images.

As part of the registration pipeline, the images were cropped and standardized to a resolution of 960x830 pixels. This normalization ensured uniformity across the dataset, facilitating consistent processing, reliable feature extraction, and accurate defect detection. Figure 4.5 illustrates the resulting images after registration.

Defects Segmentation

After obtaining the registered datasets at a standardized resolution, the next step is defect segmentation. The objective is to automatically generate masks that delineate the affected regions. For this purpose, the Segment Anything Model [95] is employed in inference mode using pre-trained weights provided by the authors. No additional training is performed,

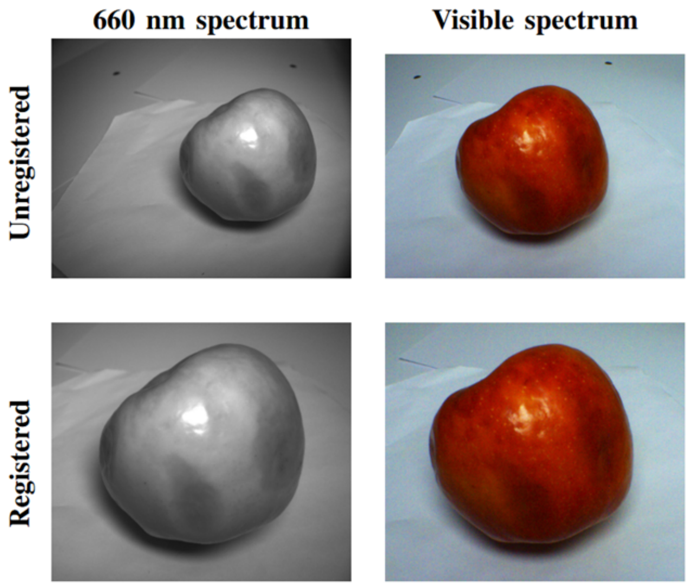


Fig. 4.5 The images in the first row show unregistered images while those in the second row show the resulting registered images.

and the parameters are adjusted to enhance sensitivity to very small regions, such as stains, ensuring the detection of subtle defect features.

The model operates exclusively in inference mode, with parameter tuning focused on accurately recognizing and delineating defects including bruises, stains, and rot. This adjustment optimizes the model's ability to capture both large and small variations in fruit surface conditions.

Following inference, post-processing techniques are applied to refine the segmentation masks. These techniques filter and retain only the relevant regions that correspond to actual defects. The post-processing step ensures high precision in defect localization, providing accurate spatial information that supports the effectiveness of subsequent classification models.

4.5.2 Spectral Subrange Selection

This study hypothesizes that focusing on a specific subrange of the visible spectrum enhances the visibility of apple defects. To evaluate this assumption, a spectral wavelength of 660 nm with a bandwidth of 60 nm is selected. This range falls within the red portion of the visible spectrum, where apples exhibit strong absorption characteristics associated with pigments such as anthocyanins and chlorophyll. Since the red spectrum is highly sensitive to physiological changes, it provides an effective basis for detecting alterations in fruit condition.

The spectral sensitivity of 660 nm is closely linked to variations in anthocyanin content, which determines apple coloration. Physiological stress, diseases, or decay modify pigment composition, producing visible alterations that serve as indicators of rot, stains, bruises, and other defects. By emphasizing this wavelength, subtle variations in apple pigmentation are captured, allowing early detection of defects that might otherwise remain imperceptible to human inspection.

Beyond its biological relevance, the 660 nm wavelength also offers practical advantages in imaging. Cameras and sensors optimized for this spectral band deliver high-resolution images with enhanced contrast, enabling precise feature extraction for defect analysis. This targeted approach increases the efficiency and reliability of classification models, reinforcing the role of spectral imaging in post-harvest quality control.

The models are trained and evaluated using two datasets: a Visible Spectrum dataset, containing images captured across the full visible range, and a 660 nm Spectrum dataset, containing images restricted to the selected wavelength. The comparison between these datasets validates whether the spectral subrange improves the performance of defect classification.

4.5.3 Training

The study employs well-established state-of-the-art object classification models, including DenseNet121, VGG19, MobileNetV1, and ResNet50, as backbones for defect detection. These architectures are widely recognized for their robustness and effectiveness in computer vision tasks. To adapt them to the classification of apple defects into three categories, namely bruises, stains, and rot, architectural refinements are introduced. A pooling layer is strategically integrated, followed by two densely connected layers that incorporate a dropout rate of 0.5 to reduce overfitting. The final layer consists of a dense output layer with three nodes, each representing the probability of belonging to a specific defect category.

The models are trained using the Adam optimizer with a learning rate of 0.0001. Pre-trained ImageNet weights are employed to initialize the backbones, enabling their use as feature extractors and enhancing the transfer of learned representations. This refined architecture, illustrated in Figure 4.6, provides a balanced framework that combines the strengths of state-of-the-art models with task-specific adaptations, ensuring robust performance in apple defect classification.

The second hypothesis proposes that incorporating segmentation masks of defects as an additional input enhances classification accuracy. To generate the segmentation masks, the Segment Anything Model [95] is employed, producing pixel-wise masks for each defect

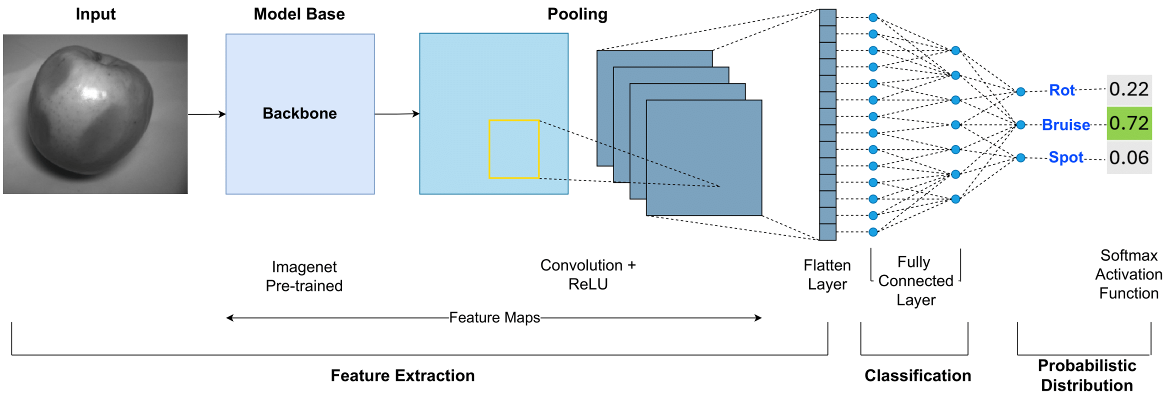


Fig. 4.6 Single-Input architecture employed for apple defect classification.

type. Using the same backbone architectures described previously (DenseNet121, VGG19, MobileNetV1, and ResNet50), a Multi-Input architecture is implemented. This network receives both the original apple image and its corresponding segmentation mask, as illustrated in Figure 4.7. Each branch processes the inputs independently through the backbone models, and the resulting feature vectors are concatenated along the depth dimension before being passed to the classifier. The Multi-Input architecture learns to associate information from the image and its mask, thereby improving the model's capacity to discriminate among defects.

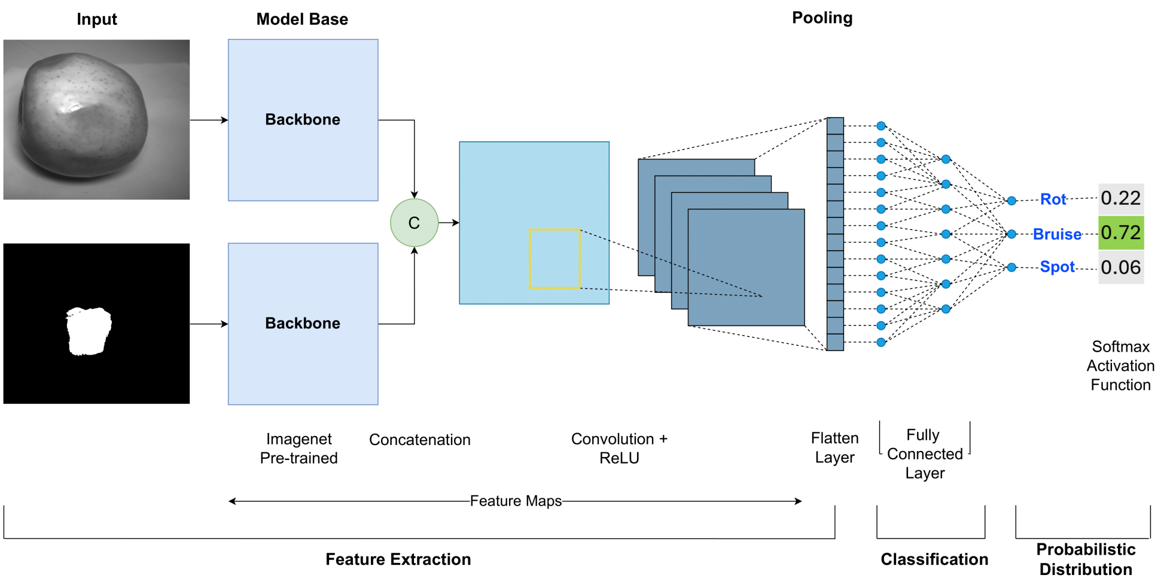


Fig. 4.7 The Multi-Input architecture implemented for apple defect classification, where both the original image and its corresponding segmentation mask are processed in parallel branches.

4.5.4 Results

To evaluate the effectiveness of deep learning models in the classification of apple defects, two complementary strategies are investigated: the Single-Input and Multi-Inputs approaches. The Single-Input method relies on a single source of spectral information, either visible spectrum or 660 nm imaging, to determine the capacity of well-established architectures to classify defects individually. In contrast, the Multi-Inputs method explores the integration of multiple sources of information, including segmentation masks and combined spectral datasets, with the aim of enhancing the discriminative power of the models. Together, these approaches provide a comprehensive assessment of how input modality influences model performance and offer valuable insights into the trade-offs between simplicity, accuracy, and computational complexity in apple defect classification.

Results with Single-Input Network Approach

The use of pre-trained ImageNet weights enables the models to converge rapidly, typically within 25 epochs. Each architecture undergoes two training sessions, one using the 660 nm spectrum dataset and the other employing the visible spectrum dataset. The classification results on the validation sets are summarized in Table 4.7. Among the tested models, MobileNetV1 achieves the highest precision, surpassing 98% in both spectral domains. Its efficiency in processing low-resolution images proves advantageous for multispectral analysis, supporting accurate apple defect classification. Notably, MobileNetV1 performs slightly better with the 660 nm dataset, reaching 98.8% precision compared to 98.26% with the visible spectrum dataset.

Table 4.7 Accuracy values of classification models on single-input architecture for 660 nm and visible spectrum datasets.

Model	660 nm Spectrum	Visible Spectrum
MobileNetV1	**98.8**	98.26
DenseNet121	95.39	98.46
ResNet50	95.79	96.10
VGG19	33.36	33.22

Results with Multi-Input Network Approach

Three experiments are conducted to evaluate different input configurations within the Multi-Inputs framework. In the first experiment, the 660 nm spectrum dataset is combined with its corresponding segmentation masks. MobileNetV1 achieves the highest precision, reaching 75.14%. In the second experiment, the visible spectrum dataset is used with its associated masks, where DenseNet121 emerges as the best-performing model, attaining 85.71% precision as shown in Table 4.8. Its dense connectivity promotes feature reuse and efficient gradient flow, enabling deep representations that enhance the classification of subtle apple defects. In the third experiment, the 660 nm spectrum dataset is paired with the visible spectrum dataset, where MobileNetV1 records the highest performance with 90.91% precision, as detailed in Table 4.9.

Table 4.8 Accuracy values of classification models on multi-input architecture for 660 nm spectrum + defect masks and visible spectrum + defect masks datasets.

Model	**660 nm spectrum + defect masks**	**Visible spectrum + defect masks**
MobileNetV1	75.14	69.85
DenseNet121	72.59	**85.71**
ResNet50	35.63	52.87
VGG19	65.78	66.04

Across all experiments, transfer learning with ImageNet pre-trained weights proves beneficial, allowing the models to capture relevant defect features effectively. Regularization techniques such as normalization and dropout further contribute to robust generalization on unseen data. When compared with the Single-Input results in Table 4.7, most models experience a decline in accuracy, except for VGG19. Remarkably, its accuracy nearly doubles when segmentation masks are incorporated, rising from approximately 33% to 65–66%. This improvement suggests that VGG19 benefits from the concatenation strategy along the depth dimension, which enhances its ability to integrate complementary features from both branches.

Although combining both spectra does not consistently improve precision compared to individual datasets, VGG19 achieves a modest 3% gain under this configuration. This

Table 4.9 Accuracy values of classification models on Multi-Input Architecture using 660 nm + visible spectrum datasets.

Model	**660 nm + visible spectrum**
MobileNetV1	**90.91**
DenseNet121	76.70
ResNet50	60.36
VGG19	36.36

outcome underscores its adaptability in leveraging multiple inputs for improved feature representation, while other models such as MobileNetV1, DenseNet121, and ResNet50 show limitations in fully exploiting the dual-input framework for this specific task.

4.5.5 Conclusions

In the Single-Input analysis, the highest-performing model achieves 98.80% precision with the 660 nm spectrum dataset, reflecting a 0.54% improvement over the visible spectrum. This outcome confirms the effectiveness of the 660 nm wavelength in highlighting apple defects such as bruises, stains, and rot. In contrast, the Multi-Inputs architecture incorporating segmentation masks generated by the SAM model does not yield improvements in precision. This limitation is likely due to inaccuracies in the segmentation masks, which introduce false positives and false negatives. Refining these masks prior to integration may enhance their utility for classification tasks.

When combining the 660 nm and visible spectrum datasets within the Multi-Inputs framework, performance surpasses that of the segmentation mask approach but remains inferior to the Single-Input scheme. These findings indicate that the integration of both spectra does not provide additional benefits for precision. Overall, the results emphasize the superiority of the Single-Input configuration with the 660 nm dataset as the most effective strategy for apple defect classification.

Chapter 5

Deform Quality

This chapter addresses the classification of fruit deformities, extending the findings reported in the publication *Fruit Deformity Classification through Single-Input and Multi-Input Architectures based on CNN Models using Real and Synthetic Images*. The proposed approach leverages convolutional neural networks trained on both real and synthetic datasets to detect morphological irregularities in fruits, an essential component of external quality assessment. By comparing Single-Input and Multi-Input architectures, the study highlights how architectural design and data diversity contribute to improving classification accuracy and robustness. The discussion presented here synthesizes those contributions and evaluates their implications for enhancing post-harvest quality control systems.

5.1 Introduction

One of the most critical stages of the post-harvest process of fruits and vegetables is quality inspection, where parameters such as ripeness, deformation, and defects are evaluated [143]. Although increasingly mechanized, this stage is still frequently performed manually by trained personnel. Its importance lies in the direct influence on the commercial value of the product, as higher quality often translates into greater market profitability. Retailers and supermarkets, regardless of scale, also rely on these parameters to guide their purchasing decisions [178]. Furthermore, consumer preferences are strongly shaped by visual attributes, reinforcing the role of appearance as a determinant factor in the acceptance of fruits and vegetables.

The quality assessment of fruits and vegetables often relies on manual inspection, making the process slow, inconsistent, and subject to the evaluators' subjective criteria. This dependence on human judgment reduces efficiency and negatively affects the commercial value of the products, with consequences for both consumers and farmers, who may face reduced profits due to inconsistent quality. To address this issue, this study proposes the use of deep learning models for the automatic classification of fruit quality, focusing specifically on the detection of deformation as a key parameter. Apples, mangoes, and strawberries are selected as case studies to evaluate the performance of the proposed approach.

5.2 Fruit Deformities

During fruit growth and development, variations in shape and size may arise from multiple factors, including genetic influences [15]. A deformed fruit presents an atypical morphology that deviates from the characteristic features of its species. This study investigates the detection of such deformities using CNNs, emphasizing the importance of accurately characterizing deformities to enable effective feature learning and robust classification.

Deformities are commonly evaluated in terms of shape and size, criteria that are standardized by international organizations responsible for quality assessment. Institutions such as the OECD [112], the USDA [76], and governmental agencies including the Republic of China [226] define and regulate these parameters. Table 5.1 summarizes the principal attributes typically assessed across different fruit species.

Table 5.1 Parameters evaluated for classifying deformities according to various international organizations.

Organizations	**Categories**	**Shape**	**Size**
OECD	Extra-Class, Class 1, Class 2	✓ Typical Shape of fruit	✗ Does not assess
USDA	Extra-Fancy, Fancy, Utility	✓ Typical Shape of fruit	✓ Measurement of the fruit's major diameter.
Republic of China	Excellent, Class 1, Class 2, Substandard	✓ Typical Shape of fruit	✓ Measurement of the fruit's major diameter. Measurement of the fruit's circumference from the top view.

Studies such as [74] and [211] evaluate apples by considering both shape and diameter. However, the use of size as a parameter requires diameter measurements, which in turn demand controlled image acquisition conditions, including a fixed working distance between the camera and the target. According to OECD standards, fruits must comply with minimum quality requirements for ripeness, defects, and shape to be classified as Extra Class, Class I, or Class II. Fruits failing to meet these standards are categorized as Ungraded and considered of lower commercial value. The following sections describe the acceptable levels of deformities in apples, mangoes, and strawberries, as defined by international standards for commercialization.

5.2.1 Apple Deformities

According to [161], the classification of apple deformities is organized into three categories: Extra Class, Class I, and Class II, as illustrated in Figure 5.1. The OECD establishes that apples must exhibit a symmetrical surface without visible deviations or irregularities, which defines the typical morphology of the fruit. During normal development, apples are expected to maintain bilateral symmetry resembling a mirror image. When this symmetry is absent, the fruit is classified as deformed [28].

Fig. 5.1 Classification of apples according to OECD standards, where deformities are determined by the presence or absence of symmetry in the fruit's morphology [17].

An apple is considered symmetrical when the distance from one side to its central axis is approximately equal to the distance from the axis to the opposite side. As illustrated in Figure 5.2, the upper portion of the apple develops uniformly, with both sides exhibiting similar growth patterns. This morphological balance is a key characteristic in quality assessment, as it defines the typical and commercially desirable shape of the fruit.

In this study, all apples are required to exhibit a symmetrical shape regardless of their variety. Consequently, apples from different varieties are incorporated into the image dataset, as their inclusion enhances the ability of convolutional neural networks to recognize and classify deformities related to symmetry.

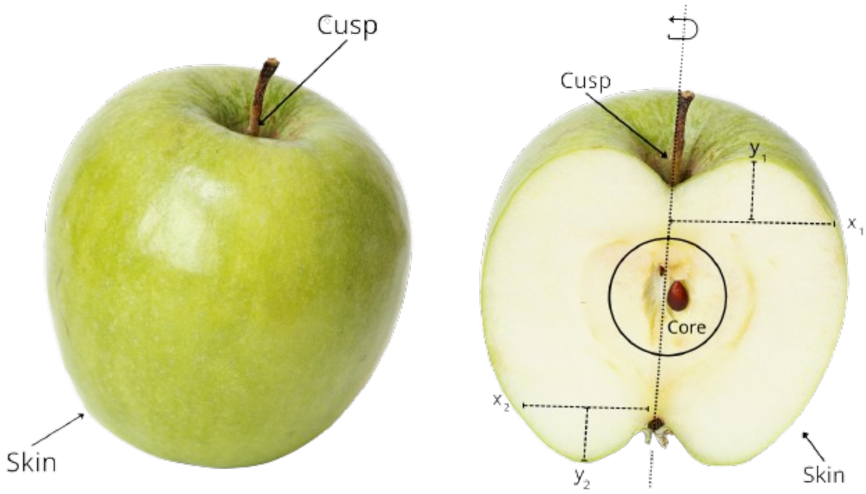


Fig. 5.2 Symmetry analysis in apples, where the distances from each side to the central axis are approximately equal, ensuring a uniform and commercially desirable shape [17].

5.2.2 Mango Deformities

According to [161], the classification of deformities in mangoes is organized into the categories shown in Figure 5.3: Extra Class, Class 1, and Class 2. The OECD specifies that mangoes should exhibit an oval shape; however, due to the diversity of mango varieties, uniformity of shape across all varieties cannot be guaranteed. This study focuses on Mangifera Indica, particularly the Tommy Atkins variety, whose characteristic ovoid shape is illustrated in Figure 5.4.

Fig. 5.3 Classification of mangoes according to the OECD standard [17].

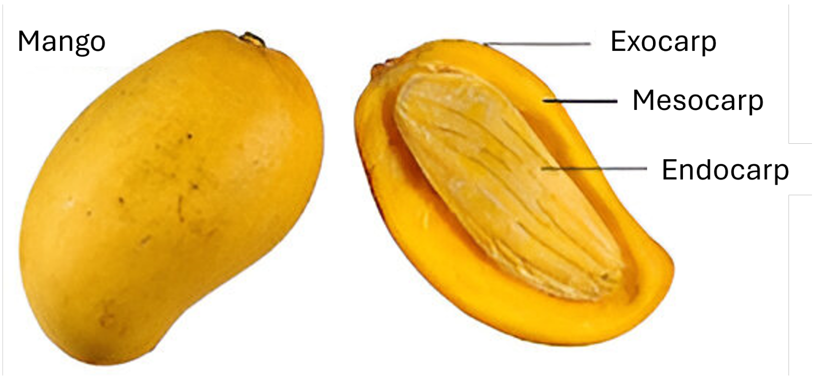


Fig. 5.4 Mango exhibiting an elongated oval shape [17].

5.2.3 Strawberry Deformities

According to [161], the classification of deformities in strawberries is organized into the categories illustrated in Figure 5.5: Extra Class, Class 1, and Class 2. As with apples and mangoes, strawberries encompass several species and varieties. However, unlike other fruits, strawberry varieties are primarily distinguished by attributes such as size and flavor, while maintaining the characteristic conical shape [116]. Consequently, all strawberry varieties are included in the dataset, as their structural form consistently supports the analysis of deformities (see Figure 5.6). A strawberry is regarded as symmetrical when its profile from the calyx to the tip resembles an inverted cone.

Fig. 5.5 Classification of strawberries according to OECD standards [17].

5.3 Dataset Generation

The effectiveness of CNN models depends heavily on the availability of extensive image datasets from which meaningful features can be extracted. Nevertheless, the mere presence

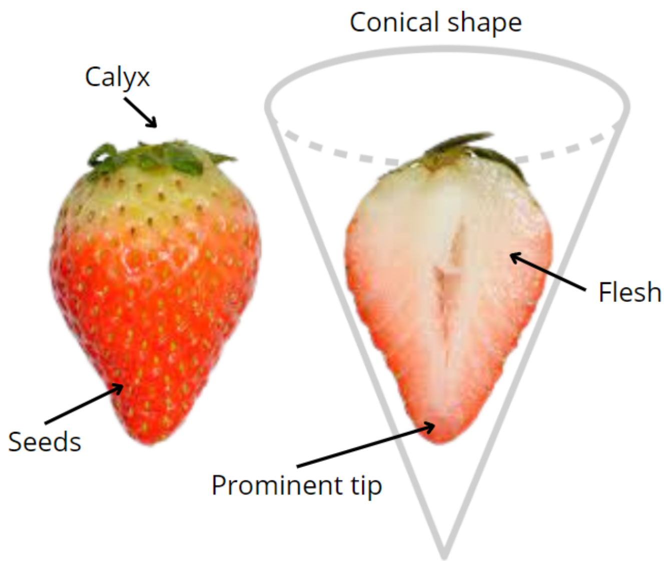


Fig. 5.6 Illustration of strawberry symmetry, showing the conical shape from calyx to tip [17].

of a large number of training images does not guarantee high accuracy, as inadequate preprocessing can significantly reduce learning efficiency [209]. This underscores the importance of curating and refining datasets to ensure the quality of input data provided to CNN models.

Several studies Garillos-Manliguez and Chiang [53], Sun et al. [188], Wang et al. [211] demonstrate the benefits of preprocessing techniques, including object segmentation, image enhancement, and the incorporation of alternative visible spectra, to increase the discriminative capacity of image datasets.

Image acquisition strategies vary depending on research goals. While public repositories are often used, strict control over acquisition parameters such as lighting and focal distance can yield superior datasets [211]. In addition, when enlarging training datasets is required, data augmentation and synthetic image generation provide valuable alternatives. For example, Chuquimarca et al. [35] employs the Unreal Engine to create realistic banana images across ripeness stages, while Pacheco et al. [147] uses DALL-E mini to generate synthetic images of apples and mangoes with defects such as rot, scabs, and blemishes. Both approaches enhance model performance by introducing variability and increasing dataset robustness.

The deformation analysis in this work is conducted exclusively using lateral views of the fruits. This acquisition strategy ensures consistent visibility of geometric deviations along the longitudinal axis, which is particularly relevant for shape-based classification. However, this approach inherently limits the assessment to side perspectives, since top or bottom views are not considered in the present study. As a result, deformation attributes that may only be distinguishable from superior viewpoints remain outside the scope of this work. Despite this constraint, the lateral orientation provides sufficiently discriminative morphological cues to enable robust categorization of deformation levels.

5.3.1 Real Image Acquisition

As with any machine learning model, CNNs require large amounts of data to effectively learn and generalize. For this study, approximately 5,000 images per deformity category, totaling 20,000 per fruit type, were considered appropriate to classify the four defined classes. To build this dataset, online repositories such as Kaggle, GitHub, and Mendeley were explored. Although no public datasets were found that matched the specific classification scheme (Extra, Class I, Class II, and Ungraded), available datasets of apples, mangoes, and strawberries were adopted and manually reclassified according to international standards. Images that did not meet the requirements, including those with complex backgrounds or taken from top or bottom perspectives, were excluded to ensure data quality. Despite this refinement, a marked imbalance among categories was observed, with the Extra Class being overrepresented. To mitigate this issue, data augmentation techniques were applied, expanding the minority classes by up to eight times and producing a balanced dataset for training.

5.3.2 Synthetic Image Generation

The generation of synthetic images is conducted in multiple stages to reach 5,000 images per class for each fruit. The process begins with Easy Diffusion version 2.5 [24], which creates diverse text-based image variants for each category. However, when the generated images either lack variation or deviate from the desired outcome, the process shifts to image-to-image generation. Following the approach in [184], small sets of images ranging from 200 to 800 are produced according to different shape scales for each class, particularly for underrepresented categories such as Class II and Ungraded. These sets are then used to train models with LORA through AUTOMATIC 1111 [72]. The trained models are subsequently applied in the Kohya generator [125], which allows the integration of learned image features with text prompts to generate additional variations. This combined strategy ensures the completion of balanced datasets, achieving 5,000 synthetic images per class for apples, mangoes, and strawberries.

5.4 Training

This section outlines the methodology applied in the development of this study and highlights its main contributions to the scientific community. Three distinct datasets are constructed: a real dataset compiled from public repositories and refined according to international standards for shape classification, a synthetic dataset generated using a combination of text-to-image and image-to-image tools to produce high-fidelity training samples, and a silhouette dataset created by extracting only fruit shapes. The complete dataset is made publicly available at https://github.com/luischuquim/Deformed-fruits.git. Using these datasets, three CNN models, VGG16, MobileNetV2, and CIDIS, are evaluated, with MobileNetV2 achieving superior performance across the experiments. The choice of classification architectures was guided by results from previous studies and by the nature of the problem addressed, which involves a limited number of categories and well-defined visual features. Given this context, the use of mature and relatively simple models was considered sufficient to capture the discriminative patterns required for accurate classification, while avoiding unnecessary computational complexity. Furthermore, the study compares two training strategies: a Single-Input architecture, which promotes progressive and linear learning, and a Multi-Input architecture, where two branches process different data modalities, RGB images and silhouette images, before feature concatenation. This comparison provides new insights into the benefits and limitations of distinct input strategies for fruit deformity classification.

The work in Hu et al. [74] focuses on identifying the shape and size of apples using a modified VGG16 model. The architecture retains the traditional two fully connected layers while incorporating four additional convolutional layers, which allows the model to reach an accuracy of 99.04%.

In Cao et al. [27], an apple size classification system is proposed using a custom CNN named LightNet. The model demonstrates strong generalization capacity, as validated with zizanias images, and achieves an accuracy of 99.31%, outperforming established models such as AlexNet, VGG11, ResNet18, MobileNet, MobileNetV2, and ShuffleNetV2.

Despite these advancements, there remains a clear scarcity of published works addressing fruit deformity classification with CNN models. This research gap underscores the need for further exploration, as deep learning approaches hold substantial potential to improve the efficiency and reliability of quality inspection in agriculture. Encouraging continued innovation in this area may lead to more accurate and automated classification systems, with direct implications for both industry practices and consumer satisfaction.

Single-Input architectures process a single type of input data at a time and represent the most common approach across a wide range of applications. In contrast, Multi-Input architectures handle two or more data modalities simultaneously, enabling the network to learn complex relationships between heterogeneous sources such as RGB images and spectral data. These architectures employ multiple branches, with each branch dedicated to processing a specific input type, before concatenating the extracted features to perform the final classification or prediction task [126, 160].

The training process begins with the CIDIS model [35], which, being a custom-designed architecture, does not include pre-trained weights. To address this, the model is first trained on the synthetic dataset to learn fundamental features, and the acquired knowledge is subsequently transferred to the real image dataset for further refinement. In parallel, VGG16 and MobileNetV2 architectures are employed under a Model Pre-training approach, where both architectures leverage pre-trained weights and are directly trained on real images. This contrasts with the CIDIS model, which undergoes a two-stage process of synthetic pre-training followed by fine-tuning with real data, thereby integrating synthetic-to-real transfer learning into the methodology.

In the implementation of a Multi-Input architecture, the primary objective is to evaluate the differences between the input images processed by each branch [126, 47, 32]. The outputs from both branches are subsequently combined through a feature fusion process, and the final classification is performed after a MLP layer. Since this approach emphasizes the analysis of fruit shape, features typically extracted by CNNs, such as color and texture, are intentionally disregarded. To address this, one branch receives RGB fruit images from both real and synthetic datasets, while the other branch is trained with corresponding silhouette images, as illustrated in Figure 5.7.

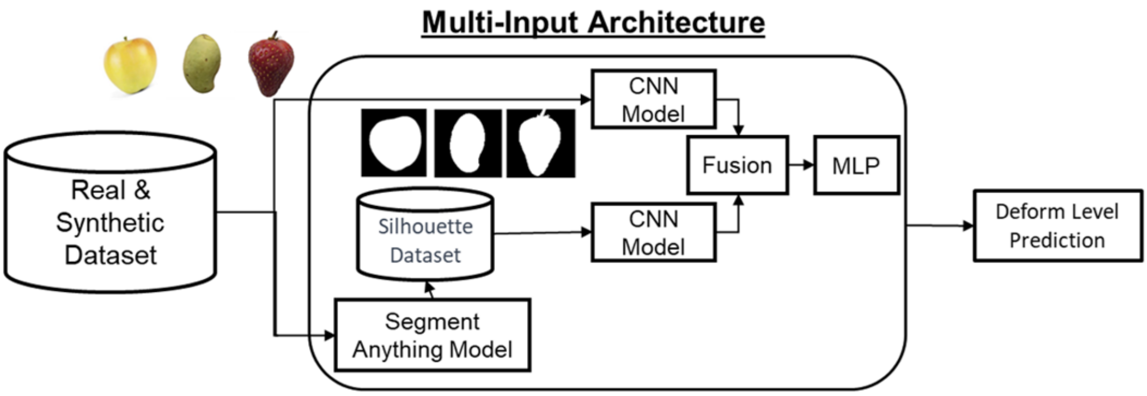


Fig. 5.7 Architecture of the Multi-Input Network, consisting of two branches: one trained with RGB images and the other with silhouette images, whose extracted features are fused before classification.

In this approach, the CIDIS, MobileNetV2, and VGG16 models are employed. The CIDIS model is trained from scratch without initial weights, whereas MobileNetV2 and

VGG16 are initialized with the pre-trained weights obtained in the Model Pre-training stage. This distinction allows for evaluating the performance of models trained with and without prior knowledge under the same experimental conditions.

5.5 Results

This section describes the real and synthetic image datasets that are acquired and generated to ensure high-quality data for evaluating CNN models, followed by a detailed presentation of the results obtained under different training approaches and model configurations. From the minimum required set of 650 real fruit images, data augmentation techniques such as rotations and horizontal and vertical flips are applied to expand the dataset and achieve the agreed split for training, validation, and testing (80/10/10). All images are standardized to a resolution of 224×224 pixels. As a result, each fruit type comprises a total of 20,000 images evenly distributed across the four deformation categories, as summarized in Table 5.2.

Given the limited availability of real images suitable for fruit deformity analysis, synthetic datasets are generated using Stable Diffusion and complementary tools. These methods produce high-quality synthetic images with dimensions of 512×512 pixels, often derived from real images to ensure consistency. Data augmentation techniques, including rotations and flips, are subsequently applied to introduce variability. As with the real dataset, a total of 20,000 images per fruit are obtained and evenly distributed across the deformity categories.

In addition, silhouette images are prepared to emphasize fruit shape, which is a key feature in deformity classification. The Segment Anything Model (SAM) is employed for the automatic extraction of silhouettes, achieving high-quality results in most cases (see Figure 5.8). When segmentation errors occur, images are either manually corrected or excluded to ensure a refined dataset, thereby supporting robust and reliable training of CNN models.

Once the datasets are prepared, hyperparameters are configured to optimize the training of each CNN model. Different combinations are evaluated, considering the impact of optimizers such as Adam, Nadam, and RMSProp, along with variable learning rates. Training sessions are typically executed for up to 30 epochs, and the final model weights are restored from the epoch that achieves the best performance, ensuring a fair and consistent comparison across models.

Table 5.2 Number of real and synthetic images in the fruit dataset.

Fruits	**Extra Class**		**First Class**		**Second Class**		**Out of Class**	
	Real	**Synthetic**	**Real**	**Synthetic**	**Real**	**Synthetic**	**Real**	**Synthetic**
Apples	4050 (*950)	5000	1741 (*3259)	5000	651 (*4349)	5000	1961 (*3039)	5000
Strawberry	2782 (*2218)	5000	875 (*4125)	5000	759 (*4241)	5000	679 (*4321)	5000
Mango	4317 (*683)	5000	1889 (*3111)	5000	655 (*4345)	5000	734 (*4266)	5000

Note: The numbers marked with (*) represent augmented images.

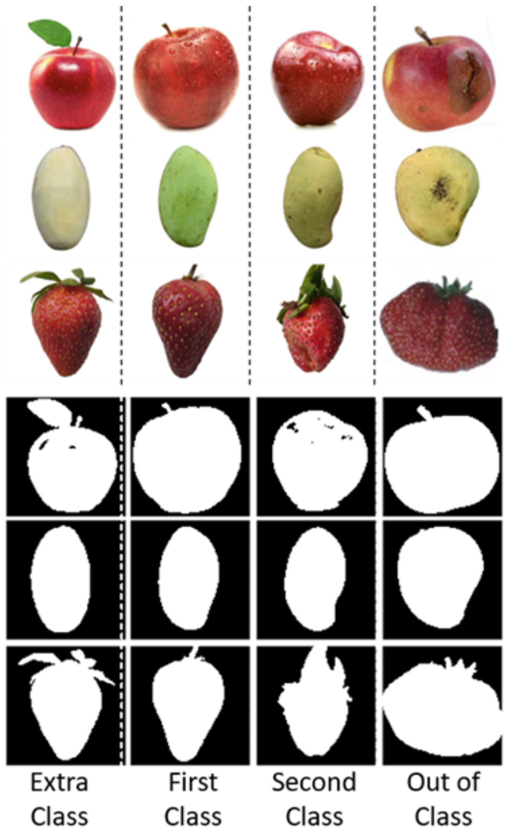


Fig. 5.8 Automatic extraction of fruit silhouettes using the SAM. The generated silhouettes are used to emphasize shape features in deformity classification tasks.

5.5.1 Results with Single-Input Network Approach

When training and evaluating the CNN models exclusively with real data, the results presented in Table 5.3 provide a comprehensive assessment of performance, including validation accuracy, validation loss, precision, recall, F1-score, and test accuracy. This table allows a clear comparison across fruits and architectures, highlighting the relative effectiveness of each approach in image classification. In contrast, the results summarized in Table 5.4, obtained through the Single-Input model (pretrained), demonstrate superior performance, with higher accuracies and lower losses. Therefore, the selection of the best-performing model is based on the combination of maximum accuracy and minimum loss achieved.

The analysis demonstrates that MobileNet consistently provides the best performance across all fruits under this training configuration. Although VGG16 achieves the highest precision for mango, its elevated loss compromises reliability, indicating that MobileNet represents a more balanced and effective architecture for accurate and stable classification outcomes.

5.5.2 Results with Multi-Input Network Approach

The methodology builds on prior studies that validate the effectiveness of Multi-Input networks in extracting complementary features from pairs of related images. In this framework, one branch of the network receives the RGB fruit image while the other processes its corresponding silhouette, enabling the model to capture similarities and differences in object shape, which is essential for accurate fruit deformity classification.

The Multi-Input network approach is applied to the complete dataset of real and synthetic images, resulting in a larger training corpus compared with the Model Pre-training strategy. This method yields consistently high performance across all tested models, with accuracies surpassing 90%. Specifically, MobileNetV2 achieves the highest accuracy for apples and strawberries, while VGG16 demonstrates superior performance for mango classification (see Table 5.5).

The results indicate that the MobileNetV2 model with the Multi-Input architecture achieves the highest performance for apples and strawberries, reaching accuracies of 92% and 93%, respectively. In contrast, the VGG16 model demonstrates superior performance for mangoes, with an accuracy of 95%. These findings confirm that the Multi-Input approach consistently improves classification outcomes compared to the Single-Input approach, as summarized in Table 5.6.

The results of this study have relevant implications for both research and the agricultural industry. The high accuracy achieved by the MobileNetV2 and VGG16 models in classifying

Table 5.3 Training Results with CIDIS, MobileNetV2, and VGG16, Model Training Approach using only Real Data.

Fruit	**Model**	**Val Accuracy**	**Val Loss**	**Precision**	**Recall**	**F1-score**	**Test Accuracy**
Apple	CIDIS	0.6815	0.8491	0.6100	0.5900	0.5500	0.5915
	MobileNetV2	**0.7848**	**0.5851**	**0.6100**	**0.6100**	**0.5700**	**0.6095**
	VGG16	0.7747	0.6073	0.5600	0.5700	0.5200	0.5685
Mango	CIDIS	0.5155	1.2260	0.4400	0.4100	0.4100	0.4400
	MobileNetV2	**0.5155**	**0.9612**	**0.6700**	**0.6700**	**0.6700**	**0.6710**
	VGG16	0.5746	1.0673	0.4900	0.4800	0.4800	0.4770
Strawberry	CIDIS	0.5601	1.3034	0.4900	0.4100	0.4000	0.4110
	MobileNetV2	**0.6971**	**1.2005**	**0.7000**	**0.7000**	**0.7000**	**0.6980**
	VGG16	0.6694	1.0141	0.6300	0.6100	0.6000	0.6100

Table 5.4 Training Results with CIDIS, MobileNetV2, and VGG16, Single-Input Network Approach.

Fruit	**Model**	**Val Accuracy**	**Val Loss**	**Precision**	**Recall**	**F1-score**	**Test Accuracy**
Apple	CIDIS	0.8956	0.3645	0.8900	0.8800	0.8800	0.8806
	MobileNetV2	**0.9325**	**0.2087**	**0.9100**	**0.9000**	**0.9000**	**0.9044**
	VGG16	0.9050	0.2511	0.9000	0.8900	0.8900	0.8931
Mango	CIDIS	0.8679	0.3819	0.8700	0.8700	0.8700	0.8694
	MobileNetV2	**0.9420**	**0.1594**	**0.9400**	**0.9400**	**0.9400**	**0.9425**
	VGG16	0.9022	0.2290	0.9600	0.9600	0.9600	0.9604
Strawberry	CIDIS	0.8010	0.5258	0.8200	0.8100	0.8200	0.8135
	MobileNetV2	**0.9189**	**0.1945**	**0.9200**	**0.9200**	**0.9200**	**0.9220**
	VGG16	0.8962	0.2646	0.9100	0.9100	0.9100	0.9050

Table 5.5 Training Results with CIDIS, MobileNetV2, and VGG16, Multi-Input Network Approach

Fruit	Model	Val Accuracy	Val Loss	Precision	Recall	F1-score	Test Accuracy
Apple	CIDIS	0.7626	1.6818	0.8639	0.8366	0.8107	0.8249
	MobileNetV2	**0.8068**	**2.2276**	**0.9257**	**0.9158**	**0.9161**	**0.9257**
	VGG16	0.8060	1.9643	0.8966	0.8920	0.8832	0.8844
Mango	CIDIS	0.9234	0.2577	0.9237	0.9232	0.9233	0.9232
	MobileNetV2	0.9440	0.1505	0.9394	0.9393	0.9390	0.9393
	VGG16	**0.9501**	**0.1696**	**0.9551**	**0.9539**	**0.9541**	**0.9551**
Strawberry	CIDIS	0.9188	0.2496	0.9145	0.9171	0.9154	0.9171
	MobileNetV2	**0.9197**	**0.2330**	**0.9455**	**0.9228**	**0.9213**	**0.9394**
	VGG16	0.8769	0.4013	0.9623	0.9204	0.9190	0.9204

Table 5.6 Results Comparison

Fruits	Models	Accuracy
Apple	MobileNetV2 (Single-Input)	0.9044
	MobileNetV2 (Multi-Input)	**0.9257**
Mango	MobileNetV2 (Single-Input)	0.9425
	VGG16 (Multi-Input)	**0.9551**
Strawberry	MobileNetV2 (Single-Input)	0.9220
	MobileNetV2 (Multi-Input)	**0.9394**

deformed fruits demonstrates the potential of these architectures for implementation in automated quality control systems. Their use reduces dependence on manual and subjective evaluations, increases consistency and speed in the inspection process, and optimizes the supply chain by ensuring that high-quality products reach consumers, ultimately enhancing profitability for farmers and suppliers. Furthermore, the demonstrated effectiveness of Multi-Input architectures compared to Single-Input provides a solid foundation for future research in the automation of quality assessment for other fruits and agricultural products.

The focus on fruit deformation strengthens and complements previous studies, such as the classification of healthy and defective fruits [147] and the analysis of banana ripeness levels [35]. These results contribute to the development of a comprehensive system for fruit categorization aligned with international standards for external quality control. In doing so, they support the establishment of consistent and reliable standards in the classification and commercialization of agricultural products.

Although the Multi-Input approach entails a higher computational cost compared to the Single-Input method, the benefits it provides in terms of precision and efficiency justify its application. By simultaneously exploiting RGB and silhouette representations, the models enhance their ability to identify deformation levels with greater accuracy. This improvement reduces classification error rates and minimizes the probability of deformed products reaching the market, mitigating negative impacts on both producers and consumers.

5.6 Conclusions

A publicly available dataset is generated, encompassing both real and synthetic images of apples, mangoes, and strawberries, thereby capturing their morphological diversity. The dataset includes 5,000 samples per fruit for each of the four predefined classes: Extra Class, First Class, Second Class, and Out of Class. In addition, 20,000 silhouette images are derived from the RGB set, offering a detailed representation of morphological structures, particularly for apples. Tools such as Stable Diffusion play a key role in synthetic data generation, facilitating the creation of a dataset robust enough to enable effective training. The results indicate that binary silhouette images enhance learning, as demonstrated by the effectiveness of the Multi-Input Network. To ensure high-quality datasets, images with consistent backgrounds and ease of segmentation are prioritized. Comparative analysis of CIDIS, MobileNetV2, and VGG16 reveals that lightweight architectures such as MobileNetV2 deliver superior performance in deformity classification.

Both the Single-Input and Multi-Input Network approaches yield strong results in deformity classification, with the latter proving more effective due to its ability to integrate complementary representations. The RGB branch captures critical features such as edges, textures, and color patterns, while the silhouette branch emphasizes contours, shapes, and pixel distributions, together enabling more precise detection of fruit deformities.

Chapter 6

Conclusions and Future Work

This chapter presents the main conclusions derived from the study and outlines potential directions for future research. The conclusions synthesize the findings of the proposed methodologies, highlighting their contributions to the advancement of automated fruit quality classification using deep learning approaches. In particular, the evaluation of real and synthetic datasets, the comparison of single-input and multi-input architectures, and the analysis of lightweight and conventional convolutional models provide a comprehensive understanding of the challenges and opportunities in addressing fruit deformity detection under variable conditions. Furthermore, the implications of these results extend beyond academic research, offering valuable insights for the agricultural industry by enabling more efficient, scalable, and objective quality control systems. Building upon these outcomes, the chapter also discusses possible future research paths, including the exploration of advanced vision transformer models, improved dataset generation techniques, and their integration into real-world industrial applications.

6.1 Conclusions

- **Create high-quality datasets comprising images of fruits to assess their external characteristics, taking into account parameters such as ripeness, deformities, and defects.** The creation of high-quality datasets stands as one of the most relevant achievements of this research. The integration of real and synthetic images partially overcame the limitation of scarce public and standardized repositories. However, the process also revealed that, although synthetic images enrich training and promote diversity, they cannot fully replicate the complexity of real-world scenarios. This underscores the need to maintain acquisition strategies that guarantee representativeness and enhance the generalization ability of the models.

- **Evaluate intelligent models for the precise classification of fruit ripeness, using relevant datasets and specific metrics to enable a detailed assessment of their maturation status, applying computational intelligence methods.** The evaluation of models for ripeness classification demonstrated the effectiveness of CNN and transformer-based architectures in identifying subtle variations in color and texture. Nevertheless, results revealed pronounced sensitivity to illumination conditions, limiting applicability in real environments. Although data augmentation improved robustness, it did not completely eliminate these vulnerabilities. These findings highlight the necessity of developing intrinsically resilient models and validation protocols that more accurately reflect the variability of industrial scenarios.

- **Evaluate intelligent models for the classification of fruit deformations, using relevant datasets and assessing their performance with appropriate metrics, incorporating computational intelligence methods.** The study of deformities showed that incorporating complementary representations, such as RGB images and silhouettes, enhances the discriminative power of models. MobileNetV2 demonstrated particularly strong performance, although results confirmed that data quality and representativeness remain critical. Deformity classification, a less explored parameter in the literature, emerges as an essential yet challenging step toward automated inspection aligned with international standards, opening a promising field of research with significant potential.

- **Evaluate intelligent models for the classification of surface defects in fruits, using relevant datasets and appropriate metrics to ensure their performance, using computational intelligence methods.** The analysis of surface defects emphasized the value of combining spectral information with multi-input architectures, which strengthened the detection of subtle patterns associated with damage and decay. Nonetheless, dependency on the quality of automatic segmentations was evident, as tools such as SAM introduced false positives and negatives that affected classification accuracy. These findings stress the importance of refining preprocessing and segmentation steps, as they represent critical factors for maximizing model performance in real post-harvest inspection applications.

List of Contributions:
This dissertation has led to the following publications, in chronological order:

- Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2023). Banana ripeness level classification using a simple CNN model trained with real and synthetic datasets. Unpublished manuscript. [35]

- Pacheco, R., González, P., Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2023). Fruit defect detection using CNN models with real and virtual data. In VISIGRAPP (VISAPP), pages 272–279. [147]

- Chuquimarca, L., Vintimilla, B., and Velastin, S. (2024). Classifying healthy and defective fruits with a multi-input architecture and CNN models. In 14th International Conference on Pattern Recognition Systems (ICPRS), pages 1–7. IEEE. [33]

- Coello, O., Coronel, M., Carpio, D., Vintimilla, B., and Chuquimarca, L. (2024). Enhancing apple's defect classification: insights from visible spectrum and narrow spectral band imaging. In 14th International Conference on Pattern Recognition Systems (ICPRS), pages 1–6. IEEE. [38]

- Beltrán, T. D., Villao, R. J., Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2024). Fruit deformity classification through single-input and multi-input architectures based on CNN models using real and synthetic images. In Iberoamerican Congress on Pattern Recognition, pages 46–62. Springer. [17]

- Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2024). A review of external quality inspection for fruit grading using CNN models. Artificial Intelligence in Agriculture, 14:1–20. Elsevier. [36]

- Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2025). Assessing deep learning model robustness for banana ripeness classification under varying illumination conditions. Smart Agricultural Technology, 101333. Elsevier. [37]

6.2 Future Work

The future research lines identified in this dissertation can be organized according to their feasibility in the short, medium, and long term. This classification provides a critical perspective on what can realistically be implemented in different academic and industrial contexts.

Short-term viability

The development of lightweight and edge-optimized models is immediately feasible, since current architectures such as MobileNet and EfficientNet already provide high accuracy with low computational cost. Their deployment on production lines represents an achievable near-term objective. Similarly, the refinement of segmentation techniques for defects and deformities, for instance improving the performance of SAM through manual corrections, can be quickly adopted. In this line, a specific project on deformity classification in bananas using 3D data is already under development, opening new possibilities for more precise morphological characterization. For deformity classification, the use of silhouettes and binary images has already proven effective, consolidating itself as a practical short-term contribution.

Medium-term viability

The creation of standardized and publicly available datasets aligned with international grading protocols is considered a medium-term challenge, as it requires coordination between academic and industrial stakeholders. The integration of spectral imaging techniques (multispectral and narrow-band) into defect and ripeness classification is also feasible within this horizon, given that the technology is available but requires investment in capture systems and calibration processes. At the algorithmic level, hybrid approaches that combine convolutional networks with transformers and self-supervised learning are promising, as they are still being optimized but hold strong potential to enhance generalization under variable conditions.

Long-term viability

The adoption of hyperspectral and 3D imaging techniques for large-scale external quality inspection is projected as a long-term objective due to the high cost of acquisition systems, calibration complexity, and heavy computational requirements. The use of advanced architectures such as transformers and graph neural networks for deformity classification also falls within this horizon, since their agricultural applications are still experimental and require significant adaptation. Finally, the construction of global frameworks for evaluation and benchmarking, including dataset quality standards, unified metrics, and interoperability protocols, demands not only technical advances but also regulatory and organizational consensus, positioning it as a transformative long-term goal.

In addition to the short, medium, and long term research directions outlined above, future work should further strengthen the pathway toward technology transfer in the food sector by expanding validation, robustness, and deployment considerations. In particular, the current approach relies predominantly on visual cues; therefore, incorporating physicochemical measurements such as °Brix, firmness, and texture would enable objective maturity verification through quantitative correlations and would improve the scientific defensibility of the proposed models. Likewise, biological variability across production batches and agricultural seasons should be explicitly modeled through broader and longitudinal data collection, domain generalization strategies, and evaluation protocols that prevent overfitting to specific lots or acquisition settings. From a regulatory standpoint, although international references such as OECD, Codex, and USDA are cited, an explicit compliance oriented validation is still required; thus, experimental comparisons against normative thresholds should be integrated, for example through comparative tables linking model outputs to standard grading criteria. Moreover, the system has not yet been assessed under realistic postharvest operating conditions, which motivates the simulation or pilot testing of an industrial workflow with real processing times, throughput constraints, and environmental variability. Finally, given the innovation oriented objective, an economic analysis estimating return on investment, considering loss reduction, labor savings, and operational efficiency, would provide actionable evidence for industrial adoption. Complementarily, future studies should clarify the most appropriate implementation points along each fruit value chain, for example processing plants for bananas and mangoes versus import and retail distribution for apples and strawberries, enabling the definition of specific user profiles such as processing industry, packing houses, supermarkets, or end consumers, thereby increasing the applicability and flexibility of the proposed quality inspection system.

References

[1] Agarwal, M., Kaliyar, R. K., and Gupta, S. K. (2022). Differential evolution based compression of cnn for apple fruit disease classification. In 2022 International Conference on Inventive Computation Technologies (ICICT), pages 76–82. IEEE.

[2] Aherwadi, N. and Mittal, U. (2022). Fruit quality identification using image processing, machine learning, and deep learning: A review. Adv. Appl. Math. Sci, 21:2645–2660.

[3] Aherwadi, N., Mittal, U., Singla, J., Jhanjhi, N., Yassine, A., and Hossain, M. S. (2022). Prediction of fruit maturity, quality, and its life using deep learning algorithms. Electronics, 11(24):4100.

[4] Akter, T., Bhattacharya, T., Kim, J.-H., Kim, M. S., Baek, I., Chan, D. E., and Cho, B.-K. (2024). A comprehensive review of external quality measurements of fruits and vegetables using nondestructive sensing technologies. Journal of Agriculture and Food Research, page 101068.

[5] Al-Dairi, M., Pathare, P. B., Al-Yahyai, R., Jayasuriya, H., and Al-Attabi, Z. (2023). Postharvest quality, technologies, and strategies to reduce losses along the supply chain of banana: A review. Trends in Food Science & Technology, 134:177–191.

[6] Arango, J. D., Staar, B., Baig, A. M., and Freitag, M. (2021). Quality control of apples by means of convolutional neural networks-comparison of bruise detection by color images and near-infrared images. Procedia CIRP, 99:290–294.

[7] Arunima, P., Gopinath, P. P., Lekshmi, P. G., and Esakkimuthu, M. (2024). Digital assessment of post-harvest nendran banana for faster grading: Cnn-based ripeness classification model. Postharvest Biology and Technology, 214:112972.

[8] Ashok, V. and Vinod, D. (2021). A novel fusion of deep learning and android application for real-time mango fruits disease detection. In Intelligent System Design, pages 781–791. Springer.

[9] Asriny, D. M., Rani, S., and Hidayatullah, A. F. (2020). Orange fruit images classification using convolutional neural networks. In IOP Conference series: Materials science and engineering, volume 803, page 012020. IOP Publishing.

[10] Avery, W., Munir, M., and Marculescu, R. (2024). Scaling graph convolutions for mobile vision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5857–5865.

[11] Azadnia, R., Fouladi, S., and Jahanbakhshi, A. (2023). Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. Results in Engineering, page 100891.

[12] Baglat, P., Hayat, A., Mendonca, F., Gupta, A., Mostafa, S. S., and Morgado-Dias, F. (2023). Non-destructive banana ripeness detection using shallow and deep learning: a systematic review. Sensors, 23(2):738.

[13] Bancal, V. and Ray, R. C. (2022). Overview of food loss and waste in fruits and vegetables: From issue to resources. In Fruits and vegetable wastes: Valorization to bioproducts and platform chemicals, pages 3–29. Springer.

[14] Barré, P., Herzog, K., Höfle, R., Hullin, M. B., Töpfer, R., and Steinhage, V. (2019). Automated phenotyping of epicuticular waxes of grapevine berries using light separation and convolutional neural networks. Computers and Electronics in Agriculture, 156:263–274.

[15] Behera, S. K., Rath, A. K., Mahapatra, A., and Sethy, P. K. (2020). Identification, classification & grading of fruits using machine learning & computer intelligence: a review. Journal of Ambient Intelligence and Humanized Computing, pages 1–11.

[16] Behera, S. K., Rath, A. K., and Sethy, P. K. (2021). Maturity status classification of papaya fruits based on machine learning and transfer learning approach. Information Processing in Agriculture, 8(2):244–250.

[17] Beltran, T. D., Villao, R. J., Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2024). Fruit deformity classification through single-input and multi-input architectures based on cnn models using real and synthetic images. In Iberoamerican Congress on Pattern Recognition, pages 46–62. Springer.

[18] Benmouna, B., Garcia-Mateos, G., Sabzi, S., Fernandez-Beltran, R., Parras-Burgos, D., and Molina-Martínez, J. M. (2022). Convolutional neural networks for estimating the ripening state of fuji apples using visible and near-infrared spectroscopy. Food and Bioprocess Technology, 15(10):2226–2236.

[19] Bhargava, A. and Bansal, A. (2021). Fruits and vegetables quality evaluation using computer vision: A review. Journal of King Saud University-Computer and Information Sciences, 33(3):243–257.

[20] Bhojanapalli, S., Chakrabarti, A., Glasner, D., Li, D., Unterthiner, T., and Veit, A. (2021). Understanding robustness of transformers for image classification. In Proceedings of the IEEE/CVF international conference on computer vision, pages 10231–10241.

[21] Biale, J. B. (1964). Growth, maturation, and senescence in fruits. Science, 146(3646):880–888.

[22] Bird, J. J., Barnes, C. M., Manso, L. J., Ekárt, A., and Faria, D. R. (2021). Fruit quality and defect image classification with conditional gan data augmentation. arXiv preprint arXiv:2104.05647.

[23] Bird, J. J., Barnes, C. M., Manso, L. J., Ekárt, A., and Faria, D. R. (2022). Fruit quality and defect image classification with conditional gan data augmentation. Scientia Horticulturae, 293:110684.

[24] Brade, S., Wang, B., Sousa, M., Oore, S., and Grossman, T. (2023). Promptify: Text-to-image generation through interactive prompt exploration with large language models. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pages 1–14.

[25] Buyukarikan, B. and Ulker, E. (2023). Classification of physiological disorders in apples using deep convolutional neural network under different lighting conditions. Multimedia Tools and Applications, pages 1–21.

[26] Caldwell, D. G. (2023). Automation in food manufacturing and processing. In Springer Handbook of Automation, pages 949–971. Springer.

[27] Cao, J., Sun, T., Zhang, W., Zhong, M., Huang, B., Zhou, G., and Chai, X. (2021). An automated zizania quality grading method based on deep classification model. Computers and Electronics in Agriculture, 183:106004.

[28] Chakrabarti, A., Michaels, T. C., Yin, S., Sun, E., and Mahadevan, L. (2021). The cusp of an apple. Nature Physics, 17(10):1125–1129.

[29] Chakraborty, S. K., Subeesh, A., Dubey, K., Jat, D., Chandel, N. S., Potdar, R., Rao, N. G., and Kumar, D. (2023). Development of an optimally designed real-time automatic citrus fruit grading–sorting machine leveraging computer vision-based adaptive deep learning model. Engineering Applications of Artificial Intelligence, 120:105826.

[30] Chang, C., Parthiban, R., Kalavally, V., Hung, Y. M., and Wang, X. (2024). Unharvested palm fruit bunch ripeness detection with hybrid color correction. Smart Agricultural Technology, 9:100643.

[31] Chen, Y., Grimplet, J., David, K., Castellarin, S. D., Terol, J., Wong, D. C., Luo, Z., Schaffer, R., Celton, J.-M., Talon, M., et al. (2018). Ethylene receptors and related proteins in climacteric and non-climacteric fruits. Plant science, 276:63–72.

[32] Choudhary, A., Mishra, R. K., Fatima, S., and Panigrahi, B. (2023). Multi-input cnn based vibro-acoustic fusion for accurate fault diagnosis of induction motor. Engineering Applications of Artificial Intelligence, 120:105872.

[33] Chuquimarca, L., Vintimilla, B., and Velastin, S. (2024a). Classifying healthy and defective fruits with a multi-input architecture and cnn models. In 2024 14th International Conference on Pattern Recognition Systems (ICPRS), pages 1–7. IEEE.

[34] Chuquimarca, L., Vintimilla, B. X., and Velastín, S. A. (2023a). Banana ripeness level classification using a simple cnn model trained with real and synthetic datasets. In VISIGRAPP.

[35] Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2023b). Banana ripeness level classification using a simple cnn model trained with real and synthetic datasets. In VISIGRAPP (5: VISAPP), pages 536–543.

[36] Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2024b). A review of external quality inspection for fruit grading using cnn models. Artificial Intelligence in Agriculture, 14:1–20.

[37] Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2025). Assessing deep learning model robustness for banana ripeness classification under varying illumination conditions. Smart Agricultural Technology, page 101333.

[38] Coello, O., Coronel, M., Carpio, D., Vintimilla, B., and Chuquimarca, L. (2024). Enhancing apple's defect classification: insights from visible spectrum and narrow spectral band imaging. In 2024 14th International Conference on Pattern Recognition Systems (ICPRS), pages 1–6. IEEE.

[39] Crane, J. H. and Gazis, R. (2020). Bacterial black spot (bbs) of mango in florida: Hs1369, 9/2020. EDIS, 2020(5).

[40] Davenel, A., Guizard, C., Labarre, T., and Sevila, F. (1988). Automatic detection of surface defects on fruit by using a vision system. Journal of Agricultural Engineering Research, 41(1):1–9.

[41] De-Arteaga, M., Herlands, W., Neill, D. B., and Dubrawski, A. (2018). Machine learning for the developing world. ACM Transactions on Management Information Systems (TMIS), 9(2):1–14.

[42] Devi, M. S., Kumar, A., Ravikumar, S., Yadav, R., and Singh, D. (2023). Thresholdedrelu orthogonal layer weight regularized densely connected convolutional networks cnn for strawberry disease prediction. In 2023 2nd International Conference for Innovation in Technology (INOCON), pages 1–5. IEEE.

[43] Devika, G., Bandaru, C. S. R., Srinivas, R. V., Girdharwal, N., Devi, M. N., et al. (2023). Real-time quality assurance of fruits and vegetables using hybrid distance based mkelm approach. In 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), pages 728–734. IEEE.

[44] Dhanush, G., Khatri, N., Kumar, S., and Shukla, P. K. (2023). A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce. Scientific African, page e01798.

[45] Dhiman, B., Kumar, Y., and Kumar, M. (2022). Fruit quality evaluation using machine learning techniques: review, motivation and future perspectives. Multimedia Tools and Applications, 81(12):16255–16277.

[46] Dikshit, A., Bartsch, A., George, A., and Farimani, A. B. (2023). Robochop: Autonomous framework for fruit and vegetable chopping leveraging foundational models. arXiv preprint arXiv:2307.13159.

[47] Dua, N., Singh, S. N., and Semwal, V. B. (2021). Multi-input cnn-gru based human activity recognition using wearable sensors. Computing, 103(7):1461–1478.

[48] Duraisamy, M., Govindaraj, R., Mohan, N., Kamalakannan, A., Bindal, S., and Titus, A. (2022). Performance evaluation of deep learning based mandarin fruit sorting system with industrial camera. In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), pages 1238–1243. IEEE.

[49] Dutta, J., Deshpande, P., and Rai, B. (2021). Ai-based soft-sensor for shelf life prediction of 'kesar'mango. SN Applied Sciences, 3(6):1–9.

[50] Enicks, D. A., Bomberger, R. A., and Amiri, A. (2020). Development of a portable lamp assay for detection of neofabraea perennans in commercial apple fruit. Plant Disease, 104(9):2346–2353.

[51] Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., Zhang, C., and Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. Journal of Food Engineering, 286:110102.

[52] Feng, L., Zhu, S., Zhou, L., Zhao, Y., Bao, Y., Zhang, C., and He, Y. (2019). Detection of subtle bruises on winter jujube using hyperspectral imaging with pixel-wise deep learning method. IEEE Access, 7:64494–64505.

[53] Garillos-Manliguez, C. A. and Chiang, J. Y. (2021). Multimodal deep learning and visible-light and hyperspectral imaging for fruit maturity estimation. Sensors, 21(4):1288.

[54] Giuggioli, N. R., Ollani, S., Zanchini, R., Danielle, B., Sparacino, A., Massaglia, S., and Merlino, V. M. (2024). The appeal of bananas: A qualitative sensory analysis and consumers' insights into tropical fruit consumption in italy. Journal of Agriculture and Food Research, 16:101110.

[55] Göbel, C., Langen, N., Blumenthal, A., Teitscheid, P., and Ritter, G. (2015). Cutting food waste through cooperation along the food supply chain. Sustainability, 7(2):1429–1445.

[56] Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.

[57] Goutam, K., Balasubramanian, S., Gera, D., and Sarma, R. R. (2020). Layerout: Freezing layers in deep neural networks. SN Computer Science, 1(5):1–9.

[58] Gunasekaran, A., Subramanian, N., and Ngai, W. T. E. (2019). Quality management in the 21st century enterprises: Research pathway towards industry 4.0.

[59] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1780–1789.

[60] Guo, X., Li, Y., and Ling, H. (2016a). Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing, 26(2):982–993.

[61] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016b). Deep learning for visual understanding: A review. Neurocomputing, 187:27–48.

[62] Gururaj, N., Vinod, V., and Vijayakumar, K. (2023). Deep grading of mangoes using convolutional neural network and computer vision. Multimedia Tools and Applications, 82(25):39525–39550.

[63] Haider, M. W., Abbas, S. M., Saeed, M. A., Farooq, U., Waseem, M., Adil, M., Javed, M. R., Haq, I. u., and Osei Tutu, C. (2025). Environmental and nutritional value of fruit and vegetable peels as animal feed: A comprehensive review. Animal Research and One Health, 3(2):149–164.

[64] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert systems with applications, 73:220–239.

[65] Hameed, K., Chai, D., and Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. Image and Vision Computing, 80:24–44.

[66] Han, J.-W., Zuo, M., Zhu, W.-Y., Zuo, J.-H., Lü, E.-L., and Yang, X.-T. (2021). A comprehensive review of cold chain logistics for fresh agricultural products: Current status, challenges, and future trends. Trends in Food Science & Technology, 109:536–551.

[67] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al. (2022). A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence, 45(1):87–110.

[68] He, H. and Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9):1263–1284.

[69] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

[70] Helwan, A., Sallam Ma'aitah, M. K., Abiyev, R. H., Uzelaltinbulat, S., and Sonyel, B. (2021). Deep learning based on residual networks for automatic sorting of bananas. Journal of Food Quality, 2021.

[71] Heo, B., Yun, S., Han, D., Chun, S., Choe, J., and Oh, S. J. (2021). Rethinking spatial dimensions of vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pages 11936–11945.

[72] Hidalgo, R., Salah, N., Chandra Jetty, R., Jetty, A., and Varde, A. S. (2023). Personalizing text-to-image diffusion models by fine-tuning classification for ai applications. In Intelligent Systems Conference, pages 642–658. Springer.

[73] Hong, G.-F., Nair, S., Lin, C.-Y., Kuan, C.-S., and Chen, S.-J. (2025). Deep learning-based detection of green-ripe pineapples via bract wilting rate analysis. Smart Agricultural Technology, page 100949.

[74] Hu, G., Zhang, E., Zhou, J., Zhao, J., Gao, Z., Sugirbay, A., Jin, H., Zhang, S., and Chen, J. (2021). Infield apple detection and grading based on multi-feature fusion. Horticulturae, 7(9):276.

[75] Huang, H., Jing, G., Wang, H., Duan, X., Qu, H., and Jiang, Y. (2014). The combined effects of phenylurea and gibberellins on quality maintenance and shelf life extension of banana fruit during storage. Scientia horticulturae, 167:36–42.

[76] Huang, K.-M., Guan, Z., and Hammami, A. (2022). The us fresh fruit and vegetable industry: An overview of production and trade. Agriculture, 12(10):1719.

[77] Ibrahim, M. S., Hamid, S. A., Muhammad, Z., Leh, N. A. M., Abdullah, S., Bakar, S. J. A., Osman, M. K., and Fadhlullah, S. Y. (2022). Cnn comparative study for apple quality classification. In 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), pages 53–58. IEEE.

[78] Irtwange, S. (2006). Maturity, quality and marketing of fruits and vegetables. Agricultural Engineering International: CIGR Journal.

[79] Ismail, A., Idris, M. Y. I., Ayub, M. N., and Por, L. Y. (2019). Investigation of fusion features for apple classification in smart manufacturing. Symmetry, 11(10):1194.

[80] Ivanovs, M., Ozols, K., Dobrajs, A., and Kadikis, R. (2022). Improving semantic segmentation of urban scenes for self-driving cars with synthetic images. Sensors, 22(6):2252.

[81] Jahanbakhshi, A., Momeny, M., Mahmoudi, M., and Zhang, Y.-D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Scientia Horticulturae, 263:109133.

[82] Jararweh, Y., Fatima, S., Jarrah, M., and AlZu'bi, S. (2023). Smart and sustainable agriculture: Fundamentals, enabling technologies, and future directions. Computers and Electrical Engineering, 110:108799.

[83] Jiang, L., Wang, Y., Yan, H., Yin, Y., and Wu, C. (2025). Strawberry fruit deformity detection and symmetry quantification using deep learning and geometric feature analysis. Horticulturae, 11(6):652.

[84] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., and Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing, 30:2340–2349.

[85] Johnson, J. M. and Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of big data, 6(1):1–54.

[86] Kader, A. A. (2002). Us grade standards. Postharvest technology of horticultural crops, 3311(287):287–300.

[87] Kanimalar, C. and Karthikeyan, M. (2024). Advances in machine learning and deep learning for automated banana disease detection: A comprehensive survey. In 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), pages 1160–1168. IEEE.

[88] Kanjilal, R., Saenz, J. E., and Uysal, I. (2025). Large-scale data-driven uniformity analysis and sensory prediction of commercial banana ripening process. Postharvest Biology and Technology, 219:113203.

[89] Karagiannis, E. (2024). Postharvest physiology of climacteric and nonclimacteric fruits and vegetables. In Oxygen, Nitrogen and Sulfur Species in Post-Harvest Physiology of Horticultural Crops, pages 1–21. Elsevier.

[90] Karakaya, D., Ulucan, O., and Turkan, M. (2019). A comparative analysis on fruit freshness classification. In 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), pages 1–4. IEEE.

[91] Karoney, E. M., Molelekoa, T., Bill, M., Siyoum, N., and Korsten, L. (2024). Global research network analysis of fresh produce postharvest technology: Innovative trends for loss reduction. Postharvest Biology and Technology, 208:112642.

[92] Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. ACM computing surveys (CSUR), 54(10s):1–41.

[93] Kikulwe, E. M., Okurut, S., Ajambo, S., Nowakunda, K., Stoian, D., and Naziri, D. (2018). Postharvest losses and their determinants: A challenge to creating a sustainable cooking banana value chain in uganda. Sustainability, 10(7):2381.

[94] Kim, W. (2022). Low-light image enhancement: A comparative review and prospects. IEEE Access, 10:84535–84557.

[95] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). Segment anything. arXiv:2304.02643.

[96] Kitinoja, L., Tokala, V. Y., and Brondy, A. (2018). Challenges and opportunities for improved postharvest loss measurements in plant-based food crops. Journal of Postharvest Technology, 6(4):16–34.

[97] Knott, M., Perez-Cruz, F., and Defraeye, T. (2023). Facilitated machine learning for image-based fruit quality assessment. Journal of Food Engineering, 345:111401.

[98] Koetter, F., Blohm, M., Drawehn, J., Kochanowski, M., Goetzer, J., Graziotin, D., and Wagner, S. (2019). Conversational agents for insurance companies: from theory to practice. In International Conference on Agents and Artificial Intelligence, pages 338–362. Springer.

[99] Kondoyanni, M., Loukatos, D., Templalexis, C., Lentzou, D., Xanthopoulos, G., and Arvanitis, K. G. (2025). Computer vision in monitoring fruit browning: Neural networks vs. stochastic modelling. Sensors, 25(8):2482.

[100] Kou, X., Feng, Y., Yuan, S., Zhao, X., Wu, C., Wang, C., and Xue, Z. (2021). Different regulatory mechanisms of plant hormones in the ripening of climacteric and non-climacteric fruits: a review. Plant Molecular Biology, 107(6):477–497.

[101] Kumar, A., Joshi, R. C., Dutta, M. K., Jonak, M., and Burget, R. (2021). Fruit-cnn: An efficient deep learning-based fruit classification and quality assessment for precision agriculture. In 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pages 60–65. IEEE.

[102] Kumari, C., Sharma, M., Kumar, V., Sharma, R., Kumar, V., Sharma, P., Kumar, P., and Irfan, M. (2022a). Genome editing technology for genetic amelioration of fruits and vegetables for alleviating post-harvest loss. Bioengineering, 9(4):176.

[103] Kumari, N., Dwivedi, R. K., Bhatt, A. K., and Belwal, R. (2022b). Automated fruit grading using optimal feature selection and hybrid classification by self-adaptive chicken swarm optimization: grading of mango. Neural computing and applications, pages 1–22.

[104] Kurama, V. (2020). A review of popular deep learning architectures: Resnet, inceptionv3, and squeezenet. Consult. August, 30.

[105] Lebaka, V. R., Wee, Y.-J., Ye, W., and Korivi, M. (2021). Nutritional composition and bioactive compounds in three different parts of mango fruit. International journal of environmental research and public health, 18(2):741.

[106] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.

[107] Lee, J.-G., Lee, Y., Kim, C. S., and Han, S. B. (2021). Codex alimentarius commission on ensuring food safety and promoting fair trade: harmonization of standards between korea and codex. Food Science and Biotechnology, 30(9):1151–1170.

[108] Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.-M., Gu, J., and Loy, C. C. (2021a). Low-light image and video enhancement using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(12):9396–9416.

[109] Li, G. and Yang, Y. (2023). Smart vision for quality apple classification using surf–harris optimizing techniques. Wireless Networks, pages 1–13.

[110] Li, L., Jia, X., and Fan, K. (2024). Recent advance in nondestructive imaging technology for detecting quality of fruits and vegetables: A review. Critical Reviews in Food Science and Nutrition, pages 1–19.

[111] Li, Y., Feng, X., Liu, Y., and Han, X. (2021b). Apple quality identification and classification by image processing based on convolutional neural networks. Scientific Reports, 11(1):16618.

[112] Lidror, A. and Prussia, S. E. (1990). Improving quality assurance techniques for producing and handling agricultural crops. Journal of Food Quality, 13(3):171–184.

[113] Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022). A survey of transformers. AI Open.

[114] Lindenberger, P., Sarlin, P.-E., and Pollefeys, M. (2023). LightGlue: Local Feature Matching at Light Speed. In ICCV.

[115] Liu, F., Snetkov, L., and Lima, D. (2017). Summary on fruit identification methods: A literature review. In 2017 3rd International Conference on Economics, Social Science, Arts, Education and Management Engineering (ESSAEME 2017). Atlantis Press.

[116] Liu, L., Ji, M.-L., Chen, M., Sun, M.-y., Fu, X.-l., Li, L., Gao, D.-S., and Zhu, C.-Y. (2016). The flavor and nutritional characteristic of four strawberry varieties cultured in soilless system. Food Science & Nutrition, 4(6):858–868.

[117] Liu, S. and Deng, W. (2015). Very deep convolutional neural network based image classification using small training sample size. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR), pages 730–734. IEEE.

[118] Liu, X., Le Bourvellec, C., Yu, J., Zhao, L., Wang, K., Tao, Y., Renard, C. M., and Hu, Z. (2022). Trends and challenges on fruit and vegetable processing: Insights into sustainable, traceable, precise, healthy, intelligent, personalized and local innovative food products. Trends in Food Science & Technology, 125:12–25.

[119] Liu, X., Wei, L., Miao, C., Zhang, Q., Yan, J., Li, S., Chen, H., and Qin, W. (2023). Application of exogenous phenolic compounds in improving postharvest fruits quality: Classification, potential biochemical mechanisms and synergistic treatment. Food Reviews International, pages 1–20.

[120] Lu, Y., Harvey, L., and Shankle, M. (2023). Survey and cost–benefit analysis of sorting technology for the sweetpotato packing lines. AgriEngineering, 5(2):941–949.

[121] Luo, W., Fan, G., Tian, P., Dong, W., Zhang, H., and Zhan, B. (2022). Spectrum classification of citrus tissues infected by fungi and multispectral image identification of early rotten oranges. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 279:121412.

[122] Lv, F., Li, Y., and Lu, F. (2021). Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision, 129(7):2175–2193.

[123] Lyu, X., Peng, W., Niu, S., Qu, Y., and Xin, Z. (2022). Evaluation of sustainable intensification of cultivated land use according to farming households' livelihood types. Ecological Indicators, 138:108848.

[124] Ma, L., Ma, T., Liu, R., Fan, X., and Luo, Z. (2022). Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5637–5646.

[125] Masrouri, M. and Qin, Z. (2024). Towards data-efficient mechanical design of bicontinuous composites using generative ai. Theoretical and Applied Mechanics Letters, 14(1):100492.

[126] Mesa, A. R. and Chiang, J. Y. (2021). Multi-input deep learning model with rgb and hyperspectral imaging for banana grading. Agriculture, 11(8):687.

[127] Mohtar, I. A., Ramli, N. S. S., and Ahmad, Z. (2019). Automatic classification of mangosteen ripening stages using deep learning. In 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), pages 44–47. IEEE.

[128] Momeny, M., Jahanbakhshi, A., Jafarnezhad, K., and Zhang, Y.-D. (2020). Accurate classification of cherry fruit using deep cnn based on hybrid pooling approach. Postharvest Biology and Technology, 166:111204.

[129] Morshed, M. S., Ahmed, S., Ahmed, T., Islam, M. U., and Rahman, A. A. (2022). Fruit quality assessment with densely connected convolutional neural network. In 2022 12th International Conference on Electrical and Computer Engineering (ICECE), pages 1–4. IEEE.

[130] Mostafidi, M., Sanjabi, M. R., Shirkhan, F., and Zahedi, M. T. (2020). A review of recent trends in the development of the microbial safety of fruits and vegetables. Trends in Food Science & Technology, 103:321–332.

[131] Naik, S. (2019). Non-destructive mango (mangifera indica l., cv. kesar) grading using convolutional neural network and support vector machine. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.

[132] Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., and Valenzuela, A. (2020). A review of convolutional neural network applied to fruit image processing. Applied Sciences, 10(10):3443.

[133] Nazir, D., Ahmed, Q. A., Khan, S., Khan, E., and Ejaz, S. (2022). Meta-analysis of machine learning methods for fruit quality prediction. Quaid-E-Awam University Research Journal of Engineering, Science & Technology, Nawabshah., 20(02):138–150.

[134] Neupane, C., Pereira, M., Koirala, A., and Walsh, K. B. (2023). Fruit sizing in orchard: A review from caliper to machine vision with deep learning. Sensors, 23(8):3868.

[135] Nguyen, H. H. C., Luong, A. T., Trinh, T. H., Ho, P. H., Meesad, P., and Nguyen, T. T. (2021). Intelligent fruit recognition system using deep learning. In International Conference on Computing and Information Technology, pages 13–22. Springer.

[136] Nguyen, T.-V.-L., Nguyen, Q.-D., and Nguyen, P.-B.-D. (2022). Drying kinetics and changes of total phenolic content, antioxidant activity and color parameters of mango and avocado pulp in refractance window drying. Polish Journal of Food and Nutrition Sciences, 72(1):27–38.

[137] Ni, J., Gao, J., Deng, L., and Han, Z. (2020). Monitoring the change process of banana freshness by googlenet. IEEE Access, 8:228369–228376.

[138] Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., and Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy: A review. Postharvest biology and technology, 46(2):99–118.

[139] Nithya, R., Santhi, B., Manikandan, R., Rahimi, M., and Gandomi, A. H. (2022). Computer vision system for mango fruit defect detection using deep convolutional neural network. Foods, 11(21):3483.

[140] Nturambirwe, J. F. I. and Opara, U. L. (2020). Machine learning applications to non-destructive defect detection in horticultural products. Biosystems Engineering, 189:60–83.

[141] Nuanmeesri, S. (2025). Enhanced hybrid attention deep learning for avocado ripeness classification on resource constrained devices. Scientific Reports, 15(1):3719.

[142] Okpala, C. O. R. and Korzeniowska, M. (2023). Understanding the relevance of quality management in agro-food product industry: From ethical considerations to assuring food hygiene quality safety standards and its associated processes. Food Reviews International, 39(4):1879–1952.

[143] Olorunfemi, B. J. and Kayode, S. E. (2021). Post-harvest loss and grain storage technology-a review. Turkish Journal of Agriculture-Food Science and Technology, 9(1):75–83.

[144] Omar, A. and MatJafri, M. (2013). Principles, methodologies and technologies of fresh fruit quality assurance. Quality Assurance and Safety of Crops & Foods, 5(3):257–271.

[145] Opara, U. L. and Ogra, I. O. (2025). An introduction to postharvest handling technology of fresh fruits and vegetables. In Sustainable Postharvest Technologies for Fruits and Vegetables, pages 3–41. CRC Press.

[146] Osei-Kwarteng, M., Ogwu, M. C., Mahunu, G. K., and Afoakwah, N. A. (2024). Postharvest food quality and safety in the global south: Sustainable management perspectives. In Food Safety and Quality in the Global South, pages 151–195. Springer.

[147] Pacheco, R., González, P., Chuquimarca, L. E., Vintimilla, B. X., and Velastin, S. A. (2023). Fruit defect detection using cnn models with real and virtual data. In VISIGRAPP (4: VISAPP), pages 272–279.

[148] Pachón-Suescún, C., Pinzón-Arenas, J., and Jiménez-Moreno, R. (2020). Fruit identification and quality detection by means of dag-cnn. Int J Adv Sci Eng Inf Technol, 10(5):2183–2188.

[149] Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., and Da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics, 10(3):279.

[150] Palakodati, S. S. S., Chirra, V. R. R., Yakobu, D., and Bulla, S. (2020). Fresh and rotten fruits classification using cnn and transfer learning. Rev. d'Intelligence Artif., 34(5):617–622.

[151] Palumbo, M., Cefola, M., Pace, B., Attolico, G., and Colelli, G. (2023). Computer vision system based on conventional imaging for non-destructively evaluating quality attributes in fresh and packaged fruit and vegetables. Postharvest Biology and Technology, 200:112332.

[152] Pandey, V. K., Srivastava, S., Dash, K. K., Singh, R., Mukarram, S. A., Kovács, B., and Harsányi, E. (2023). Machine learning algorithms and fundamentals as emerging safety tools in preservation of fruits and vegetables: a review. Processes, 11(6):1720.

[153] Pardede, J., Sitohang, B., Akbar, S., and Khodra, M. L. (2021). Implementation of transfer learning using vgg16 on fruit ripeness detection. International Journal of Intelligent Systems & Applications, 13(2).

[154] Pathmanaban, P., Gnanavel, B., Anandan, S. S., and Sathiyamurthy, S. (2023). Advancing post-harvest fruit handling through ai-based thermal imaging: applications, challenges, and future trends. Discover Food, 3(1):27.

[155] Patra, A., Abdullah, S., and Pradhan, R. C. (2022). Review on the extraction of bioactive compounds and characterization of fruit industry by-products. Bioresources and Bioprocessing, 9(1):14.

[156] Payasi, A. and Sanwal, G. (2010). Ripening of climacteric fruits and their control. Journal of food Biochemistry, 34(4):679–710.

[157] Perez, R. M., Cheein, F. A., and Rosell-Polo, J. R. (2017). Flexible system of multiple rgb-d sensors for measuring and classifying fruits in agri-food industry. Computers and Electronics in Agriculture, 139:231–242.

[158] Pérez-Llorca, M., Muñoz, P., Müller, M., and Munné-Bosch, S. (2019). Biosynthesis, metabolism and function of auxin, salicylic acid and melatonin in climacteric and non-climacteric fruits. Frontiers in plant science, 10:136.

[159] Phasinam, K., Kassanuk, T., and Shabaz, M. (2022). Applicability of internet of things in smart farming. Journal of Food Quality, 2022(1):7692922.

[160] Pipitsunthonsan, P., Pan, L., Peng, S., Khaorapapong, T., Nakasathien, S., Channumsin, S., and Chongcheawchamnan, M. (2023). Palm bunch grading technique using a multi-input and multi-label convolutional neural network. Computers and Electronics in Agriculture, 210:107864.

[161] Porat, R., Lichter, A., Terry, L. A., Harker, R., and Buzby, J. (2018). Postharvest losses of fruit and vegetables during retail and in consumers' homes: Quantifications, causes, and means of prevention. Postharvest biology and technology, 139:135–149.

[162] Pourdarbani, R., Sabzi, S., Dehghankar, M., Rohban, M. H., and Arribas, J. I. (2023). Examination of lemon bruising using different cnn-based classifiers and local spectral-spatial hyperspectral imaging. Algorithms, 16(2):113.

[163] Prabhu, A., Sangeetha, K., Likhitha, S., and Lakshmi, S. S. (2021). Applications of computer vision for defect detection in fruits: A review. In 2021 International Conference on Intelligent Technologies (CONIT), pages 1–10. IEEE.

[164] Pu, Y.-Y., Feng, Y.-Z., and Sun, D.-W. (2015). Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review. Comprehensive Reviews in Food Science and Food Safety, 14(2):176–188.

[165] Qian, J., Ruiz-Garcia, L., Fan, B., Villalba, J. I. R., McCarthy, U., Zhang, B., Yu, Q., and Wu, W. (2020). Food traceability system from governmental, corporate, and consumer perspectives in the european union and china: A comparative review. Trends in Food Science & Technology, 99:402–412.

[166] Rajini, M. and Persis, V. (2025). Developing an iot and ml-driven platform for fruit ripeness evaluation and spoilage detection: A case study on bananas. e-Prime-Advances in Electrical Engineering, Electronics and Energy, page 100896.

[167] Ramadhan, Y. A., Djamal, E. C., Kasyidi, F., and Bon, A. T. (2020). Identification of cavendish banana maturity using convolutional neural networks. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Dubai, United Arab Emirates, pages 10–12.

[168] Ramos, R. P., Gomes, J. S., Prates, R. M., Simas Filho, E. F., Teruel, B. J., and dos Santos Costa, D. (2021). Non-invasive setup for grape maturation classification using deep learning. Journal of the Science of Food and Agriculture, 101(5):2042–2051.

[169] Rashmi, H. and Negi, P. (2022). Chemistry and physiology of fruits and vegetables. In Advances in Food Chemistry: Food Components, Processing and Preservation, pages 439–470. Springer.

[170] Reid, M. S. (2002). Maturation and maturity indices. Postharvest technology of horticultural crops, 3:55–62.

[171] Rivero Mesa, A. and Chiang, J. (2021). Non-invasive grading system for banana tiers using rgb imaging and deep learning. In 2021 7th International Conference on Computing and Artificial Intelligence, pages 113–118.

[172] Rodriguez, M., Pastor, F., and Ugarte, W. (2021). Classification of fruit ripeness grades using a convolutional neural network and data augmentation. In 2021 28th Conference of Open Innovations Association (FRUCT), pages 374–380. IEEE.

[173] Saeed, F., Afzaal, M., Hussain, M., and Tufail, T. (2021). Advances in assessing product quality. In Food losses, sustainable postharvest and food technologies, pages 191–218. Elsevier.

[174] Safari, Y., Nakatumba-Nabende, J., Nakasi, R., and Nakibuule, R. (2024). A review on automated detection and assessment of fruit damage using machine learning. IEEE access, 12:21358–21381.

[175] Sanaeifar, A., Bakhshipour, A., and De La Guardia, M. (2016). Prediction of banana quality indices from color features using support vector regression. Talanta, 148:54–61.

[176] Saragih, R. E. and Emanuel, A. W. (2021). Banana ripeness classification based on deep learning using convolutional neural network. In 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), pages 85–89. IEEE.

[177] Saranya, N., Srinivasan, K., and Kumar, S. P. (2022). Banana ripeness stage identification: a deep learning approach. Journal of Ambient Intelligence and Humanized Computing, 13(8):4033–4039.

[178] Shewfelt, R. L. and Prussia, S. E. (2022). Challenges in handling fresh fruits and vegetables. In Postharvest Handling, pages 167–186. Elsevier.

[179] Shokravi, Z., Shokravi, H., Atabani, A., Lau, W. J., Chyuan, O. H., and Ismail, A. F. (2022). Impacts of the harvesting process on microalgae fatty acid profiles and lipid yields: Implications for biodiesel production. Renewable and Sustainable Energy Reviews, 161:112410.

[180] Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1):1–48.

[181] Singh, A., Vaidya, G., Jagota, V., Darko, D. A., Agarwal, R. K., Debnath, S., and Potrich, E. (2022). Recent advancement in postharvest loss mitigation and quality management of fruits and vegetables using machine learning frameworks. Journal of Food Quality, 2022(1):6447282.

[182] Singh, D., Sharma, R. R., and Kesharwani, A. K. (2021). Postharvest losses of horticultural produce. In Postharvest handling and diseases of horticultural produce, pages 1–24. CRC Press.

[183] Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4):427–437.

[184] Somepalli, G., Singla, V., Goldblum, M., Geiping, J., and Goldstein, T. (2023). Diffusion art or digital forgery? investigating data replication in diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6048–6058.

[185] SS, V. C., Hareendran, A., and Albaaji, G. F. (2024). Precision farming for sustainability: An agricultural intelligence model. Computers and Electronics in Agriculture, 226:109386.

[186] Starobin, S. M. (2021). Credibility beyond compliance: Uncertified smallholders in sustainable food systems. Ecological Economics, 180:106767.

[187] Sumathi, K. and Vinod, V. (2022). Classification of fruits ripeness using cnn with multivariate analysis by sgd. Neural Network World, 32(6):319.

[188] Sun, L., Liang, K., Song, Y., and Wang, Y. (2021). An improved cnn-based apple appearance quality classification method with small samples. IEEE Access, 9:68054–68065.

[189] Swords, D. S., Bednarski, B. K., Messick, C. A., Tillman, M. M., Chang, G. J., and You, Y. N. (2022). Quality and location of the surgical episode mediate a large proportion of socioeconomic-based survival disparities in patients with resected stage i–iii colon cancer. Annals of surgical oncology, 29(1):706–716.

[190] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence, volume 31.

[191] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9.

[192] Tan, M. and Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In International conference on machine learning, pages 10096–10106. PMLR.

[193] Tan, W., Zhao, C., and Wu, H. (2016). Intelligent alerting for fruit-melon lesion image based on momentum deep learning. Multimedia Tools and Applications, 75(24):16741–16761.

[194] Taormina, P. J. (2021). Purposes and principles of shelf life determination. Food Safety and Quality-Based Shelf Life of Perishable Foods, pages 1–26.

[195] Tasara, B. T. and Qomariyah, N. N. (2022). Studying the performance of transfer learning on cnn models for fruit sorting. In 2022 International Conference on ICT for Smart Society (ICISS), pages 01–06. IEEE.

[196] Tian, S. and Xu, H. (2022). Nondestructive methods for the quality assessment of fruits and vegetables considering their physical and biological variability. Food Engineering Reviews, 14(3):380–407.

[197] Toivonen, P. M., Mitcham, E. J., and Terry, L. A. (2014). Postharvest care and the treatment of fruits and vegetables. Horticulture: Plants for People and Places, Volume 1: Production Horticulture, pages 465–483.

[198] Tonucci, D. (2023). A historical overview of food regulations in the united states. In History of food and nutrition toxicology, pages 183–214. Elsevier.

[199] Tripathi, M. K. and Maktedar, D. D. (2020). A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey. Information Processing in Agriculture, 7(2):183–203.

[200] Tripathi, M. K. and Maktedar, D. D. (2021). Optimized deep learning model for mango grading: Hybridizing lion plus firefly algorithm. IET Image Processing.

[201] Ukwuoma, C. C., Zhiguang, Q., Bin Heyat, M. B., Ali, L., Almaspoor, Z., and Monday, H. N. (2022). Recent advancements in fruit detection and classification using deep learning techniques. Mathematical Problems in Engineering, 2022:1–29.

[202] Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M.-F., and Debeir, O. (2011). Automatic grading of bi-colored apples by multispectral machine vision. Computers and electronics in agriculture, 75(1):204–212.

[203] Upadhyay, N. and Bhargava, A. (2025). Artificial intelligence in agriculture: applications, approaches, and adversities across pre-harvesting, harvesting, and post-harvesting phases. Iran Journal of Computer Science, pages 1–24.

[204] Urugo, M. M., Yohannis, E., Teka, T. A., Gemede, H. F., Tola, Y. B., Forsido, S. F., Tessema, A., Suraj, M., and Abdu, J. (2024). Addressing post-harvest losses through agro-processing for sustainable development in ethiopia. Journal of Agriculture and Food Research, page 101316.

[205] Vasumathi, M. and Kamarasan, M. (2021). An effective pomegranate fruit classification based on cnn-lstm deep learning models. Indian Journal of Science and Technology, 14(16):1310–1319.

[206] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[207] Velesaca, H. O., Suárez, P. L., Mira, R., and Sappa, A. D. (2021). Computer vision based food grain classification: A comprehensive survey. Computers and Electronics in Agriculture, 187:106287.

[208] Vicente, A. R., Manganaris, G. A., Darre, M., Ortiz, C. M., Sozzi, G. O., and Crisosto, C. H. (2022). Compositional determinants of fruit and vegetable quality and nutritional value. In Postharvest handling, pages 565–619. Elsevier.

[209] Vujović, Ž. et al. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6):599–606.

[210] Wang, F., Zheng, J., Tian, X., Wang, J., Niu, L., and Feng, W. (2018). An automatic sorting system for fresh white button mushrooms based on image processing. Computers and electronics in agriculture, 151:416–425.

[211] Wang, J., Huo, Y., Wang, Y., Zhao, H., Li, K., Liu, L., and Shi, Y. (2022a). Grading detection of "red fuji" apple in luochuan based on machine vision and near-infrared spectroscopy. Plos one, 17(8):e0271352.

[212] Wang, J., Tan, W., Niu, X., and Yan, B. (2019). Rdgan: Retinex decomposition based adversarial learning for low-light enhancement. In 2019 IEEE international conference on multimedia and expo (ICME), pages 1186–1191. IEEE.

[213] Wang, W., Wu, X., Yuan, X., and Gao, Z. (2020). An experiment-based review of low-light image enhancement methods. Ieee Access, 8:87884–87917.

[214] Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.-P., and Kot, A. (2022b). Low-light image enhancement with normalizing flow. In Proceedings of the AAAI conference on artificial intelligence, volume 36, pages 2604–2612.

[215] Wu, A., Zhu, J., and Ren, T. (2020). Detection of apple defect using laser-induced light backscattering imaging and convolutional neural network. Computers & Electrical Engineering, 81:106454.

[216] Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., and Zhang, L. (2021). Cvt: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pages 22–31.

[217] Xiang, Q., Wang, X., Li, R., Zhang, G., Lai, J., and Hu, Q. (2019). Fruit image classification based on mobilenetv2 with transfer learning technique. In Proceedings of the 3rd international conference on computer science and application engineering, pages 1–7.

[218] Yang, L., Cui, B., Wu, J., Xiao, X., Luo, Y., Peng, Q., and Zhang, Y. (2024). Automatic detection of banana maturity—application of image recognition in agricultural production. Processes, 12(4):799.

[219] Yang, S., Zhou, D., Cao, J., and Guo, Y. (2023). Lightingnet: An integrated learning method for low-light image enhancement. IEEE Transactions on Computational Imaging, 9:29–42.

[220] Zabalza, J., Ren, J., Zheng, J., Han, J., Zhao, H., Li, S., and Marshall, S. (2015). Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. IEEE transactions on geoscience and remote sensing, 53(8):4418–4433.

[221] Zeid, M. A.-E., El-Bahnasy, K., and Abo-Youssef, S. (2021). Multiclass colorectal cancer histology images classification using vision transformers. In 2021 tenth international conference on intelligent computing and information systems (ICICIS), pages 224–230. IEEE.

[222] Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision, pages 818–833. Springer.

[223] Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., and Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. Food Research International, 62:326–343.

[224] Zhang, Y., Lian, J., Fan, M., and Zheng, Y. (2018). Deep indicator for fine-grained classification of banana's ripening stages. EURASIP Journal on Image and Video Processing, 2018(1):1–10.

[225] Zhao, S., Hao, S., Zi, B., Xu, H., and Wong, K.-Y. K. (2024). Bridging different language models and generative vision models for text-to-image generation. In European Conference on Computer Vision, pages 70–86. Springer.

[226] Zhou, J.-h., Kai, L., and Liang, Q. (2015). Food safety controls in different governance structures in china's vegetable and fruit industry. Journal of Integrative Agriculture, 14(11):2189–2202.

[227] Zhu, L. and Spachos, P. (2021). Support vector machine and yolo for a mobile food grading system. Internet of Things, 13:100359.

[228] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1):43–76.

CERTIFICACIÓN DE APROBACIÓN DE PROYECTO DE TITULACIÓN

Región del Maule, 19 de enero de 2026

Ph. D.

Angel Sappa

Coordinador(a) del Programa

Doctorado en Ingeniería Eléctrica

En su Despacho

De mi consideración:

Yo, **Marco Mora**, de nacionalidad chilena, portador del pasaporte No. F474322375, en mi calidad de Evaluador del Proyecto de Titulación correspondiente al **Doctorado en Ingeniería Eléctrica con mención en Control y Automatización**, **Segunda** Cohorte, de la Escuela Superior Politécnica del Litoral (ESPOL), certifico lo siguiente:

Con fecha **30/09/2025**, acepté ser revisor del estudiante **Luis Enrique Chuquimarca Jiménez,** con C.I. **1104610132**, para el desarrollo del proyecto de titulación denominado: *External Quality Inspection of Post-harvest Fruits using Computational Intelligence Methods"*.

Certifico que este trabajo de titulación fue supervisado de manera continua durante todo su desarrollo, revisado en cada una de sus etapas y, finalmente, aprobado por mi persona en su versión final, entregada el día [19/01/2026].

Particular que pongo en su conocimiento para los fines pertinentes.

Atentamente,

Powered by ecert

Firma electrónica avanzada
MARCO ANTONIO MORA COFRE
2026.01.19 12:48:13 -0300

Marco Mora
Evaluador del Proyecto
Correo electrónico: **marcomoracofre@gmail.com**

Nota: Adjunto evidencia de la validez de la firma electrónica

Evidencia: