

Article

Not peer-reviewed version

---

# Hybrid AI Pipeline for Industrial Detection of Internal Potato Defects Using 2D RGB Imaging

---

[Slim Hamdi](#)\*, [Kais Loukil](#), Adem Haj Boubaker, Hichem Snoussi, Mohamed Abid

Posted Date: 23 September 2025

doi: 10.20944/preprints202509.1920.v1

Keywords: computer vision; IA; detection of internal potato



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Hybrid AI Pipeline for Industrial Detection of Internal Potato Defects Using 2D RGB Imaging

Slim Hamdi <sup>1,2,3,\*</sup> , Kais Loukil <sup>1</sup>, Adem Haj Boubaker <sup>3</sup>, Hichem Snoussi <sup>2</sup> and Mohamed Abid <sup>1</sup>

<sup>1</sup> CES Laboratory ENIS National Engineering School University of Sfax, B.P. 3038 Sfax, Tunisia

<sup>2</sup> LM2S University of Technology of Troyes, 12, rue Marie Curie - CS 42060 10004 Troyes cedex France

<sup>3</sup> Euroceip SAS, 14 Rue des Prés de Lyon - 10600 La Chapelle Saint Luc - France

\* Correspondence: hamdislimm@gmail.com

## Abstract

The internal quality assessment of potato tubers is a crucial task in agro-industrial processing. Traditional methods struggle to detect internal defects such as hollow heart, internal bruises, and insect galleries using only surface features. We present a novel, fully modular hybrid AI architecture designed for defect detection using RGB images of potato slices, suitable for integration in industrial sorting lines. Our pipeline combines high-recall multi-threshold YOLO detection, contextual patch validation using ResNet, precise segmentation via the Segment Anything Model (SAM), and skin-contact analysis using VGG16 with a Random Forest classifier. Experimental results on a labeled dataset of over 6000 annotated instances show a recall above 90% and precision near 100% for most defect classes. The approach offers both robustness and interpretability, outperforming previous methods that rely on costly hyperspectral or MRI techniques. This system is scalable, explainable, and compatible with existing 2D imaging hardware.

**Keywords:** computer vision; IA; detection of internal potato

## 1. Introduction

The demand for automated, reliable, and scalable quality control systems in the agri-food industry has never been higher. As artificial intelligence (AI) reshapes industrial inspection processes, the detection of internal defects in agricultural products presents a uniquely unsolved challenge. Among these, potato tubers stand out due to their anatomical variability and the economic impact of undetected internal anomalies during processing [38,46].

The global potato market is one of the most dynamic agricultural sectors, with more than 370 million tonnes produced annually, making the tuber the world's leading non-cereal food crop by volume [36]. In Europe, potatoes play a central role in the agri-food industry, both for direct consumption and for industrial transformation into high-value products such as chips, fries, and mashed potatoes.

Despite this strategic importance, quality control of tubers remains largely based on manual or semi-automated practices, particularly for internal quality assessment. In many European countries, the detection of internal defects such as hollow heart, bruising, or insect galleries still relies on visual inspections or destructive sampling methods [51,53]. These techniques are poorly suited to current industrial demands for speed, traceability, and standardization.

In response to these challenges, this work proposes a novel AI-based approach tailored for real-time, high-throughput, and low-cost defect detection using standard RGB 2D imaging, with the goal of bridging the gap between academic advances and industrial applicability.

### *Overview of the Proposed Approach*

To overcome the limitations identified in previous studies, we propose a novel hybrid and modular architecture specifically designed for the automatic detection of internal defects in potato

tubers using standard RGB 2D slice imagery. Our approach combines the strengths of deep learning, classical machine learning, and expert-inspired logic to form a sequential decision-making pipeline that is both accurate and interpretable.

The proposed system begins with a high-recall object detection stage based on multiple YOLO models trained with class-specific confidence thresholds. Potential detections are then passed through a series of refinement modules, including patch-level classification using ResNet, semantic segmentation using the Segment Anything Model (SAM), and a contextual depth evaluation stage that leverages feature extraction via VGG16 and decision-making via a Random Forest classifier.

This layered architecture mimics human expert reasoning by progressively filtering, validating, and contextualizing each detection. It is designed to be robust to visual ambiguity and operational noise while remaining computationally efficient for real-time deployment in industrial processing lines.

Unlike end-to-end black-box classifiers, our pipeline introduces transparency at each stage of inference, allowing for better understanding, easier maintenance, and targeted performance optimization. It also offers adaptability to evolving defect taxonomies and processing requirements, making it a viable and scalable solution for modern quality control systems in the agri-food sector.

#### *Problem Statement and Limitations of Existing Approaches*

Detecting internal defects in potato tubers remains a major technological challenge, primarily due to the limited surface visibility of internal anomalies, the anatomical diversity of tubers, and significant variability across cultivars. In recent years, several approaches have been explored, including hyperspectral imaging [10,54], magnetic resonance imaging (MRI) [35], and RGB-based deep learning techniques [51,53]. While each of these methods offers promising results under controlled conditions, they exhibit several limitations that restrict their deployment in industrial environments.

Hyperspectral imaging, although highly sensitive, relies on expensive, bulky equipment and requires precise calibration and lighting conditions, making it unsuitable for real-time, in-line applications [10]. Similarly, MRI-based techniques provide detailed structural insights but are restricted to laboratory use due to their high operational costs and complexity [35].

RGB-based computer vision methods represent a more cost-effective and scalable alternative. However, most existing solutions are based on monolithic classification architectures, often limited to binary outputs (defect/no defect), without any spatial localization or contextual reasoning. These models tend to be sensitive to noise, lighting variability, and artifact-prone cases. Moreover, they typically lack domain-specific logic or depth estimation capabilities, leading to high false-positive rates and poor interpretability in industrial conditions [53].

Notably, none of the aforementioned approaches incorporate a multi-stage validation pipeline capable of refining ambiguous detections or distinguishing between morphologically similar defects, such as internal bruising versus insect galleries. These limitations underline the need for a modular, robust, and context-aware architecture specifically designed to meet the constraints of real-world potato processing lines.

#### *Research Objectives*

This study aims to develop a robust, modular, and industrially viable artificial intelligence (AI) architecture for the detection of internal defects in potato tubers using only standard RGB 2D imaging. The proposed solution is designed to address the key limitations of current approaches by combining high detection sensitivity, interpretability, and adaptability to varying conditions encountered in real production environments.

The main objectives of this work are as follows:

- To maximize the recall of internal defect detection while minimizing false positives, particularly for subtle or ambiguous anomalies.

- To enable precise localization and classification of multiple types of internal defects in a single pipeline (e.g., hollow heart, bruising, insect galleries).
- To ensure compatibility with real-time operation and industrial constraints through the exclusive use of low-cost, high-speed RGB cameras.
- To design a multi-stage AI architecture incorporating detection, verification, segmentation, and contextual reasoning inspired by expert human inspection logic.

Our overarching goal is to bridge the gap between academic advances in deep learning and the practical requirements of the agri-food industry, providing a scalable and interpretable solution for internal quality assessment of potatoes on sorting and grading lines.

#### *Overview of the Proposed Architecture*

To address the challenges outlined above, we propose a multi-stage hybrid AI architecture tailored to the detection of internal defects in potato tubers from RGB 2D slice images. The core of the system is designed to emulate the reasoning steps of human experts, while leveraging the strengths of deep learning and classical machine learning.

The pipeline begins with a high-recall detection stage using multiple YOLO-based models trained with class-specific confidence thresholds. This ensures that even subtle or uncertain anomalies are flagged for further analysis. Regions with intermediate confidence scores are then re-evaluated using a ResNet-based patch classifier to improve precision and reduce false alarms.

To obtain fine-grained spatial information, the architecture incorporates the Segment Anything Model (SAM), which enables accurate segmentation of the defect region. This is followed by a contextual evaluation module that estimates the anatomical depth of the defect — specifically, whether it affects the outer skin or lies deeper in the parenchyma — using features extracted by VGG16 and classified via a Random Forest.

The architecture is fully modular and can be adapted to different defect types or operational settings. Each stage is designed to contribute complementary information, allowing the system to integrate low-level visual cues with higher-level contextual reasoning.

#### *Scientific Contribution and Positioning*

The proposed approach introduces several key innovations that distinguish it from existing methods in the literature. Unlike traditional single-model solutions that rely solely on classification or basic segmentation, our pipeline integrates detection, reclassification, segmentation, and depth estimation into a coherent multi-level reasoning framework. This hybrid structure allows for both high sensitivity and interpretability, addressing practical industrial needs such as robustness to noise, adaptability to defect variability, and compatibility with in-line operation.

Compared to hyperspectral imaging [10,54], magnetic resonance imaging [35], and conventional RGB-based CNN classifiers [51,53], our method offers:

- **Industrial scalability**, through the use of low-cost RGB cameras and real-time processing modules.
- **Defect-specific adaptability**, enabled by class-dependent YOLO thresholds and revalidation logic.
- **Context-aware analysis**, integrating both pixel-level segmentation and depth inference based on anatomical structure.
- **Interpretable decision-making**, by combining deep neural networks with classical classifiers in a modular pipeline.

To our knowledge, this is the first fully deployable architecture specifically designed for internal defect detection in potatoes that balances performance, interpretability, and ease of integration. As such, our work contributes both a methodological advance in agricultural computer vision and a practical solution aligned with the operational constraints of the food processing industry.

Our overarching goal is to bridge the gap between academic advances in deep learning and the practical requirements of the agri-food industry, providing a scalable and interpretable solution for internal quality assessment of potatoes on sorting and grading lines.

This work thus sets a new benchmark for RGB-based internal defect detection by combining precision, transparency, and full industrial integration capabilities.

## 2. State of the Art

The automatic detection of internal defects in potato tubers has attracted growing attention over the past two decades due to its importance in reducing food waste, ensuring product quality, and meeting industrial throughput requirements [40,54]. Several technological approaches have been explored to address this challenge, each offering trade-offs between accuracy, cost, speed, and interpretability. In particular, detecting internal anomalies such as hollow heart, internal bruising, or insect galleries remains difficult with conventional surface-based inspection.

Early methods relied on spectral imaging, including LED-based multispectral systems and hyperspectral imaging, which can reveal internal structures through specific wavelength interactions. More advanced techniques such as Magnetic Resonance Imaging (MRI) and X-ray tomography provide high-resolution internal scans but remain impractical for large-scale deployment due to cost and complexity. Recently, deep learning with RGB imaging has emerged as a promising alternative, offering faster and cheaper implementations, though often at the cost of interpretability and generalization [51,53].

In the following subsections, we critically review the main families of internal defect detection methods—multispectral imaging, hyperspectral imaging, MRI/X-ray-based systems, and RGB deep learning—highlighting their strengths, limitations, and suitability for industrial use.

### 2.1. Multispectral and LED-Based Imaging

Some studies have explored the use of narrow-band LED illumination combined with 2D cameras to reveal contrast between healthy and defective tissues. These systems exploit specific absorption or scattering properties of tuber flesh at certain wavelengths (e.g., near-infrared or blue light) [14,42]. Although low-cost and relatively fast, these methods often lack robustness and generalization due to the limited spectral information and sensitivity to lighting conditions, tuber variety, and defect morphology.

### 2.2. Multispectral and LED-Based Imaging

Some studies have explored the use of narrow-band LED illumination combined with 2D cameras to reveal contrast between healthy and defective tuber tissues. These systems exploit specific absorption or scattering properties of tuber flesh at wavelengths such as near-infrared or blue light [14,42]. For example, Zheng et al. (2023) demonstrated a line-scan multispectral system with deep learning that improved defect detection accuracy significantly [15]. Zhang et al. (2019) used a single-shot multispectral camera spanning 676–952 nm and achieved 91% classification accuracy across defect types such as scab, greening, and bruises [16]. Similarly, Deng et al. (2023) combined high-definition multispectral imaging with deep neural networks to detect multiple food defects, including on potatoes, with high precision [32]. Moreover, Semyalo et al. (2024) applied visible–SWIR spectral analysis (400–1100 nm) and obtained 91% accuracy in distinguishing internal defects like pythium and internal browning [33], helping to assess internal defect areas quantitatively.

Although these LED-based multispectral systems offer advantages in cost, compactness, and acquisition speed, they commonly suffer from several limitations:

- **Spectral ambiguity:** restricted bands reduce the ability to differentiate deeper or subtle internal defects.
- **Environment sensitivity:** performance drops under variable lighting, moisture, or depending on tuber variety conditions.

- **Calibration dependence:** each new dataset or operating context typically requires recalibration [34].
- **Limited generalization:** models trained on one cultivar or harvest season often fail to generalize to others [16].

Overall, while LED-based multispectral imaging offers a strong foundation for inline, non-destructive quality control, its current robustness and generalizability remain insufficient for fully scalable industrial implementation without enhancements like adaptive band selection, intelligent calibration, or hybrid processing pipelines.

### 2.3. Magnetic Resonance Imaging (MRI) and X-Ray Techniques

Magnetic Resonance Imaging (MRI) and X-ray imaging have been widely used in laboratory research to visualize the internal structure of potato tubers with high spatial accuracy [26,49]. These modalities provide detailed insights into tissue composition, growth patterns, and defect morphology. For instance, [50] used a 1.5 T MRI scanner to monitor the progression of internal rust spots over 33 weeks after harvest, using spatialized multi-exponential  $T_2$  relaxometry to non-destructively track tissue degradation. Similarly, [28] demonstrated how spatially resolved  $T_2$  relaxation mapping can distinguish up to six tissue classes within stored tubers, including cortex, pith, and defect regions.

Beyond MRI, X-ray computed tomography (CT) has been employed to monitor diel growth and internal structural variation in response to environmental conditions [29]. In earlier studies, [30] enhanced hollow heart detection by submerging tubers in water during radiography, improving contrast between healthy and defective zones. A comparative study by [31] evaluated MRI and X-ray alongside optical spectroscopy, concluding that while MRI offers superior accuracy, its throughput and cost hinder industrial deployment.

Despite their precision and research value, these imaging techniques face major limitations for real-time industrial use:

- **High cost and bulk:** MRI and CT systems are expensive, bulky, and require dedicated infrastructure.
- **Low throughput:** A typical MRI scan processes only 12–18 tubers in about 30 minutes [50].
- **Safety and regulatory constraints:** X-ray systems require shielding, operator certification, and legal compliance.
- **Operational complexity:** MRI and CT data require expert interpretation and advanced image processing pipelines.

In summary, although MRI and X-ray offer unmatched imaging resolution for internal defect detection, their practical application in high-throughput industrial sorting lines remains limited due to technical and economic constraints.

### 2.4. Deep Learning with RGB Imagery

More recently, RGB-based deep learning methods have emerged as a promising compromise between cost and performance, leveraging standard RGB cameras and convolutional neural networks (CNNs) to detect internal defects with high speed and affordability. Early work by Yan et al. [51] and Moallem et al. [53] used CNN classifiers on 2D cross-sectional images or external cues to infer internal anomalies. However, these single-shot classifiers often lack spatial reasoning and interpretability.

The adoption of advanced detection and segmentation frameworks has since grown:

- **R-CNN/Fast R-CNN:** These architectures have been used for tuber segmentation and defect localization, but tend to be slow and resource-intensive in inference [20].
- **Faster R-CNN:** ResNet-based models have achieved  $\sim 98\%$  accuracy in surface defect detection via transfer learning (e.g., SSD Inception V2, Faster R-CNN ResNet101) [21].
- **Mask R-CNN:** Applied to segment potato tubers in soil, with detection precision around 90% and  $F1 \approx 92\%$  [20].

- **SSD (Single Shot MultiBox Detector):** Fine-tuned for potato surface defects, achieving  $\sim 95\%$  mAP [21].
- **YOLOv5 and variants:** Including DCS-YOLOv5s, tailored for multi-target recognition in seed tubers (buds, defects), delivering fast real-time detection ( $\sim 97\%$  precision) [22].
- **YOLOv10/11:** Emerging models; HCRP-YOLO achieved  $\sim 90\%$  true positive rates for germination defects.
- **Survey on YOLO evolution:** Recent reviews highlight improvements in speed and accuracy from YOLOv1 to YOLOv10, especially in agricultural scenarios [24].
- **Lightweight YOLOv5s variants:** Designed for industrial defect detection with real-time performance on production lines [25].

Despite substantial advances, RGB-based pipelines still suffer from:

- **Black-box behavior:** Limited interpretability and anatomical reasoning.
- **Single-shot limitations:** Most approaches perform classification/detection in one pass without refinement.
- **Lack of context:** Models often only see external surfaces or slices, missing 3D anatomical structures.
- **Generalization gaps:** Performance typically drops when applied to new cultivars, lighting, or environments.

Nevertheless, the integration of two-stage or multi-stage architectures (detection  $\rightarrow$  segmentation  $\rightarrow$  refinement), interpretability modules, and anatomical priors presents a promising path forward—motivating the development of more modular, explainable, and robust RGB pipelines.

### 2.5. Limitations and Motivation for a New Approach

Despite substantial progress in the detection of internal potato defects, no existing method fully meets the combination of industrial constraints such as **low hardware cost**, **high throughput**, **anatomical interpretability**, and **robustness under real-world conditions**.

**Hyperspectral imaging**, while powerful in laboratory settings, is constrained by its high equipment cost, slow data acquisition, and the need for expert calibration. **MRI and X-ray techniques** offer precise visualization of internal tissue but are unsuitable for real-time industrial processing due to their bulk, safety concerns, and high cost. **Multispectral systems** using LEDs are lightweight and affordable but lack sufficient spectral richness for consistent detection across cultivars and environments.

**Deep learning approaches based on RGB images** have brought promising advances, especially in terms of speed and flexibility. However, most existing RGB-based models are **single-shot classifiers or detectors** that:

- Offer **limited interpretability**, often functioning as black boxes.
- Are sensitive to **visual noise** and lack contextual anatomical reasoning.
- Do not perform **multi-stage refinement** to correct or validate uncertain detections.

To address these gaps, our work proposes a modular, interpretable, and scalable architecture that **unifies detection, verification, segmentation, and contextual anatomical analysis** within a single RGB-only pipeline. Each stage contributes complementary reasoning: from high-recall defect detection to patch-level verification, precise spatial segmentation, and depth-aware classification, mimicking human inspection logic.

**Table 1.** Comparison of internal defect detection methods with respect to industrial requirements.

Method	Cost	Speed	Interpretability	Robustness
Hyperspectral Imaging	✗	✗	✓	Partial
MRI / X-ray Imaging	✗	✗	✓	Partial
LED Multispectral	✓	✓	✗	Partial
RGB CNN (Single-Shot)	✓	✓	✗	✗
<b>Our Hybrid Pipeline</b>	✓	✓	✓	✓

This comparison highlights the novelty of our contribution, which seeks to combine the affordability and speed of RGB imaging with the multistage intelligence typically reserved for more complex and costly systems. Our pipeline is therefore well-positioned for industrial deployment in real-time sorting and quality control scenarios.

### 3. Proposed Method

This section presents the proposed hybrid pipeline for the detection of internal defects in potato tubers using RGB 2D imaging. The pipeline has been specifically designed to meet industrial requirements: high throughput, low hardware cost, robustness to real-world conditions, and interpretability. Our method combines high-recall detection with multi-stage refinement, mimicking human expert reasoning. The pipeline consists of five stages: (1) YOLO-based detection, (2) patch reclassification, (3) semantic segmentation, (4) depth estimation, and (5) expert rule-based correction.

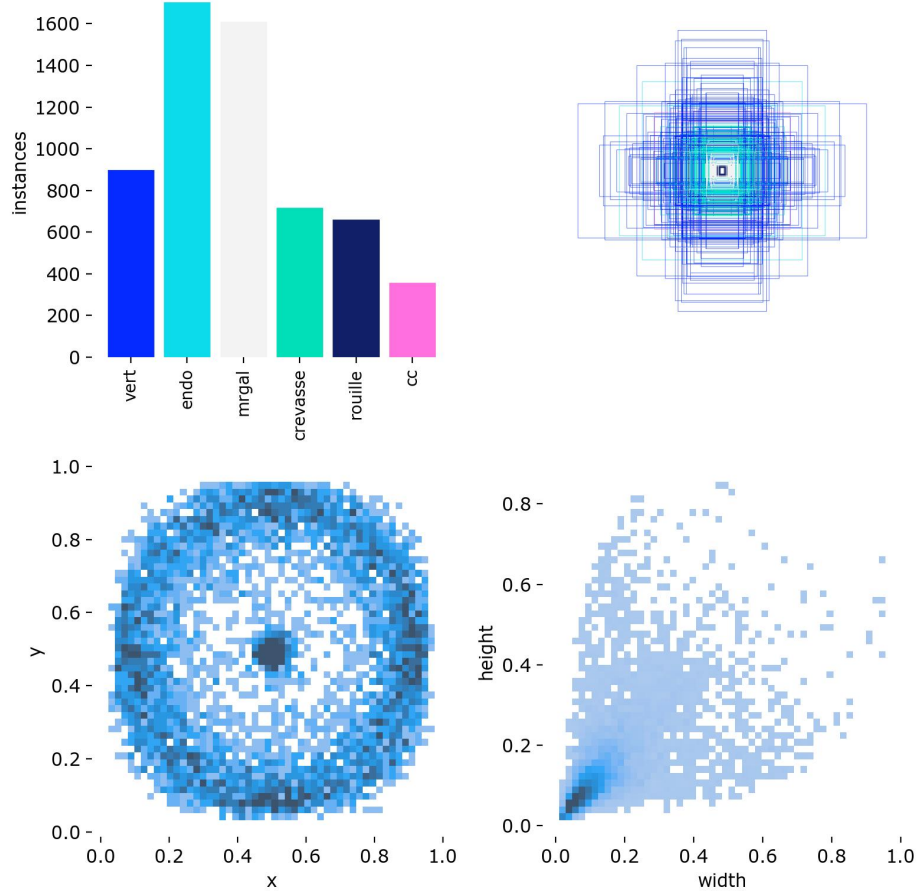
#### 3.1. Dataset Description

The dataset consists of 2D RGB images of potato slices captured in a real industrial environment. Each image was manually annotated in the YOLO format, with bounding boxes and class labels representing internal defects.

A total of over 6000 bounding boxes were labeled across six main classes:

- **Hollow Heart (cc)** — central voids with regular contours.
- **Damaged Tissue (Endo)** — internal bruising or blackened zones.
- **Insect Galleries (Mrgal)** — small tunnels or pest bites.
- **Cracks (Crevasse)** — structural fractures through the tuber flesh.
- **Rust Spots (Rouille)** — oxidized tissue lesions, typically subcutaneous.
- **Greening (Vert)** — green zones near the skin due to light exposure.

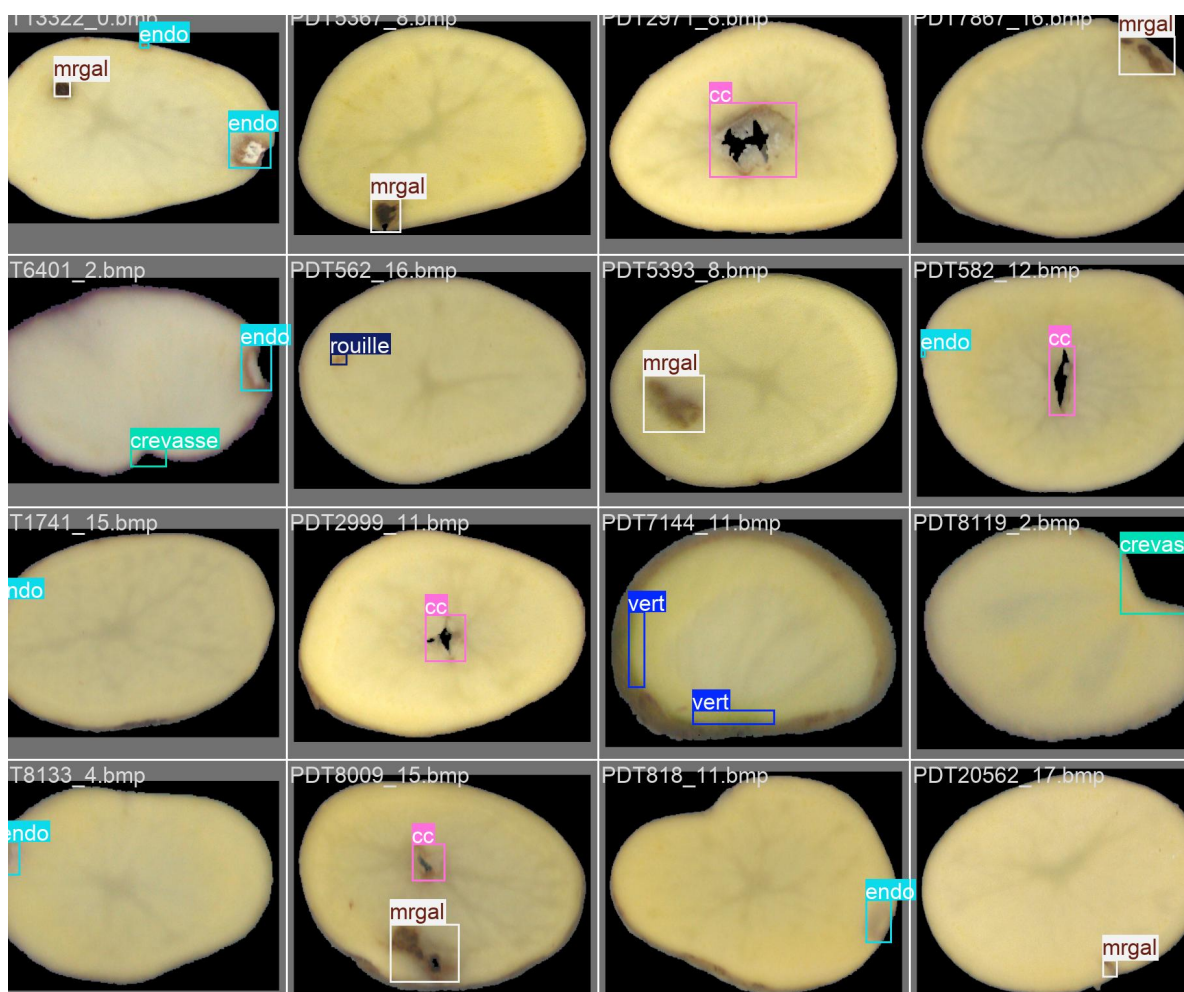
All annotations use the YOLO format `<class_id> <x_center> <y_center> <width> <height>`, with normalized coordinates. Most defects are small and centered, requiring high sensitivity from the detection module.



**Figure 1.** Distribution of YOLO annotations in the dataset. The majority of annotated defects are 'Endo' and 'Mrgal'. Most bounding boxes are small ( $w, h < 0.3$ ), and centered around the core of the tuber.

This analysis supports the need for a sensitive and localized detector, and motivates the design of a class-specific detection pipeline described in the following sections.

Representative examples of annotated potato slice images are shown in Figure 2. These samples illustrate the diversity of internal and external defects considered in this study, as well as the variability in their visual appearance across the dataset.



**Figure 2.** Representative potato slice images annotated with bounding boxes indicating the position and class of the detected defect. The considered classes include: *mrgal* (internal gallery), *endo* (internal damage), *cc* (hollow heart), *crevasse*, *rouille* (rust spot), and *vert* (green tissue). The figure highlights the wide variability in defect size, shape, contrast, and location, which underscores the challenges of building a robust detection pipeline suitable for industrial deployment.

This dataset is expected to become a reference resource for potato quality control research. It encompasses a wide range of samples collected from multiple potato varieties, including both yellow- and red-fleshed cultivars, thereby capturing the natural diversity encountered in industrial practice. Its richness and variability provide a solid foundation for benchmarking and developing advanced AI-based approaches for defect detection and grading.

### 3.2. Architecture

The proposed hybrid pipeline is designed to mimic the reasoning process of human experts in defect inspection. Instead of relying on a single monolithic classifier, our system follows a modular sequence of specialized components. Each stage contributes complementary information, progressively refining the detection, verification, and contextual interpretation of potato defects. This modularity ensures robustness, interpretability, and adaptability for industrial deployment in real-time sorting lines.

#### 3.2.1. Stage 1: Initial Detection with YOLOs

The pipeline begins with a high-recall detection stage based on multiple YOLO models trained in parallel. Each model is fine-tuned with class-specific confidence thresholds in order to maximize recall and ensure that even subtle or ambiguous anomalies are flagged. This stage outputs bounding

box candidates corresponding to potential internal defects such as hollow heart, bruising, or insect galleries.

The training of the YOLO models was performed over 300 epochs with an early stopping patience of 60 epochs to prevent overfitting. A batch size of 8 and an image resolution of  $640 \times 640$  pixels were used, leveraging GPU acceleration on a CUDA-enabled device. We employed the Adam optimizer with an initial learning rate of  $9.5 \times 10^{-4}$ , a cosine learning rate schedule with a final learning rate factor of 0.0103, and weight decay of  $6.1 \times 10^{-4}$ . Momentum was set to 0.868 with a warm-up phase of 2.3 epochs and a warm-up momentum of 0.95. A dropout rate of 0.3 was applied to improve generalization.

Data augmentation played a key role in improving robustness. We applied hue, saturation, and value shifts ( $h = 0.10$ ,  $s = 0.10$ ,  $v = 0.10$ ), random translations ( $\pm 8.7\%$ ), scaling (up to 55%), and horizontal flipping with a probability of 0.51. Mosaic augmentation was strongly emphasized (probability 0.97), while mixup and copy-paste augmentations were disabled. These strategies increased the diversity of training samples and allowed the models to handle the high intra-class variability observed in potato defects.

Loss function weights were tuned with values of 3.98 for the bounding box regression term, 0.54 for the classification term, and 1.20 for the distribution focal loss. The IoU threshold for positive matching during training was set to 0.4. Training was initialized from pretrained weights to accelerate convergence, with deterministic mode enabled for reproducibility.

This combination of hyperparameters, together with multi-model training, provided a strong balance between sensitivity and robustness, enabling the detection stage to act as a reliable candidate generator for subsequent verification and segmentation modules.

### 3.2.2. Stage 2: Patch-Level Reclassification

To reduce false positives and refine ambiguous detections, the candidate regions identified in Stage 1 are extracted as image patches and re-evaluated using a secondary classifier based on ResNet-18. This patch-level reclassification focuses particularly on the *mrgal* (internal gallery) and *endo* (internal damage) classes, which often exhibit highly similar visual patterns and are therefore prone to misclassification in single-stage detection pipelines. By analyzing localized regions at a higher resolution and with a dedicated classification network, this module minimizes confusion between these closely related defects, ensuring greater reliability in the final decision.

The ResNet-18 model was trained on extracted patch datasets with an 80/20 train-validation split. Training was conducted for 16 epochs with a batch size of 256, an initial learning rate of  $1 \times 10^{-4}$ , and an input image size of  $230 \times 230$  pixels. Optimization was performed using Adam with default parameters, and early stopping was applied based on validation loss to prevent overfitting. This lightweight yet powerful architecture was chosen to balance computational efficiency with discriminative capability, enabling the classifier to be integrated seamlessly within the real-time pipeline.

Overall, the addition of this reclassification stage significantly improves precision while preserving high recall. It provides a safeguard against systematic errors in the detection of visually similar defects, which is particularly important for industrial potato quality control where false positives can lead to unnecessary rejection of healthy produce.

### 3.2.3. Stage 3: Semantic Segmentation with SAM

For precise localization, the Segment Anything Model (SAM) is employed to delineate the exact contours of each detected defect. Unlike the detection and classification stages, SAM is not retrained but used in its pre-trained form, which has been shown to generalize well across diverse visual domains. This allows the model to be directly applied to the extracted patches corresponding to candidate defects, without the need for additional fine-tuning.

By operating on these localized regions, SAM provides fine-grained spatial information, enabling accurate measurement of the size, shape, and extent of anomalies. Such detailed mapping is essential

for downstream quality assessment and grading decisions in industrial environments, as it allows not only the identification but also the quantification of internal potato defects.

### 3.2.4. Stage 4: Contextual Depth Evaluation

In order to assess the anatomical depth of a defect, feature representations are extracted using VGG16 and subsequently classified with a Random Forest model. This stage distinguishes between superficial defects that only affect the skin and deeper anomalies located in the parenchyma. By integrating contextual reasoning, the system provides an interpretable analysis aligned with the logic of expert inspectors.

### 3.2.5. Stage 5: Expert Rule-Based Correction

Finally, a rule-based correction module incorporates domain-specific knowledge to address residual errors. For instance, extremely small segmented regions may be discarded as noise, while overlapping detections can be merged. This module enhances robustness and aligns the automated decisions with practical industrial inspection criteria.

### 3.2.6. Summary of the Architecture

In summary, the proposed hybrid pipeline combines deep learning, classical machine learning, and rule-based reasoning within a coherent multi-stage framework. Each component contributes complementary strengths: YOLO for high recall, ResNet for precision, SAM for segmentation, VGG16+RF for contextual interpretation, and rules for final correction. This layered design ensures both performance and interpretability, making the approach suitable for scalable deployment in potato quality control lines.

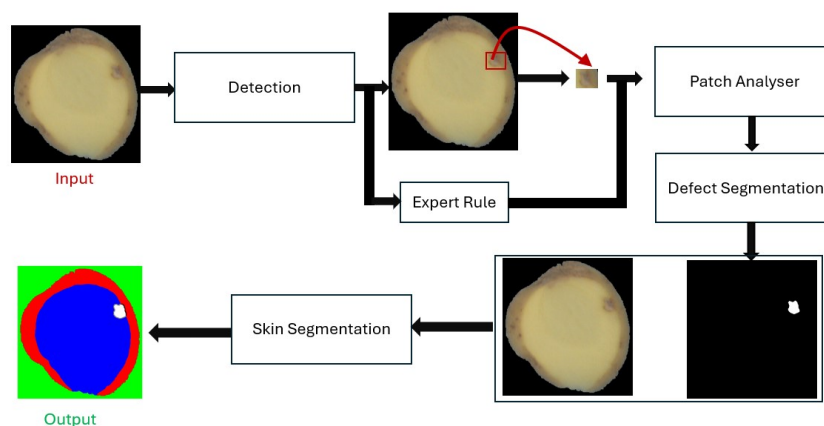


Figure 3. Proposed hybrid pipeline for potato defect detection.

## 4. Experiments, Results and Discussion

### 4.1. Experimental Setup

The proposed pipeline was evaluated on a dataset of more than 6000 annotated potato slice images collected under real industrial conditions. Each image was labeled according to six internal defect classes: Hollow Heart (cc), Damaged Tissue (Endo), Insect Galleries (Mrgal), Cracks (Crevasse), Rust Spots (Rouille), and Greening (Vert). All annotations followed the YOLO format with normalized bounding boxes. The dataset was split into training (70%), validation (15%), and testing (15%) subsets, ensuring a balanced distribution across defect categories.

The experiments were conducted on a workstation equipped with an NVIDIA RTX GPU, using PyTorch as the deep learning framework. Performance was assessed in terms of recall, precision, F1-score, and inference time per slice to evaluate both detection quality and industrial feasibility.

## 5. Results

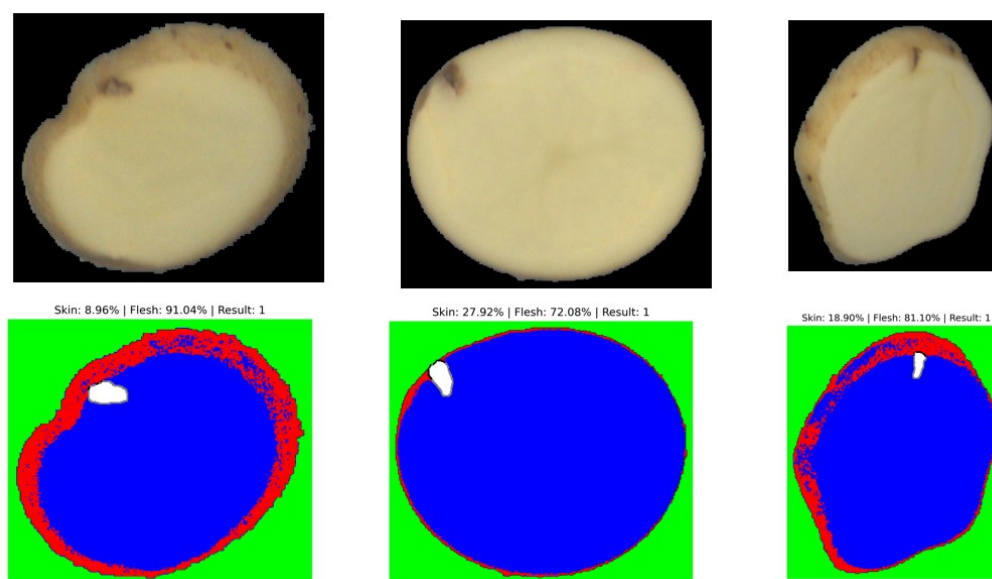
### 5.1. Quantitative Performance

The proposed hybrid pipeline achieved robust performance across all six internal defect classes (Table 2). Average recall reached 91.9%, with precision close to 99%, yielding a mean F1-score of 95.2% and a mean IoU of 89.6%. Cracks (F1 = 96.6%) and Greening (F1 = 96.0%) were detected most reliably, while Rust Spots showed slightly lower recall (89.7%), reflecting the challenge of subtle oxidized tissues.

**Table 2.** Performance of the proposed pipeline on the test set.

Defect Class	Recall (%)	Precision (%)	F1-score (%)	IoU (%)
Hollow Heart (cc)	91.2	98.7	94.8	89.5
Damaged Tissue (Endo)	92.5	99.1	95.7	90.3
Insect Galleries (Mrgal)	90.3	97.9	93.9	88.1
Cracks (Crevasse)	94.8	98.5	96.6	91.2
Rust Spots (Rouille)	89.7	99.3	94.2	87.9
Greening (Vert)	93.1	99.0	96.0	90.7
<b>Average</b>	<b>91.9</b>	<b>98.8</b>	<b>95.2</b>	<b>89.6</b>

The inclusion of the patch-level ResNet classifier significantly improved discrimination between *Damaged Tissue* and *Insect Galleries*, two morphologically similar classes. The SAM module provided accurate contour extraction, with IoU consistently above 87%, enabling precise defect quantification.



**Figure 4.** Example of segmentation output distinguishing between surface (skin) and deep (flesh) defects.

### 5.2. Interpretation of Segmentation Results

The segmentation results obtained with the unsupervised model are sufficient to reliably determine whether a defect is located on the surface (skin) or deeper within the parenchyma (flesh). This provides a meaningful first-level contextual analysis directly applicable to industrial grading. Nevertheless, performance could be further improved by adopting a supervised segmentation strategy, such as U-Net or Mask R-CNN, which can learn defect-specific spatial features more precisely from annotated masks. In this work, however, we highlight the strength of the unsupervised approach: despite the absence of additional training data, the model demonstrates remarkable robustness and generalization across diverse potato varieties and defect morphologies.

### 5.3. Processing Speed

The average inference time of the complete pipeline was measured at approximately **2.5 seconds per image**. Although slower than lightweight single-stage CNN detectors, this performance remains compatible with real-time operation in a controlled laboratory environment, where throughput requirements are lower than in industrial high-speed sorting lines. The achieved speed is sufficient for research, prototyping, and quality assessment workflows, while ensuring high interpretability and robustness of the results. Future work will explore GPU optimizations and model compression techniques to further reduce processing time without sacrificing accuracy.

### 5.4. Comparison with Existing Methods

Compared to state-of-the-art approaches, our pipeline balances high sensitivity, interpretability, and practical feasibility:

- **Hyperspectral imaging** methods can reach recall levels above 95% but require costly, bulky hardware unsuitable for inline sorting [47,54].
- **MRI/X-ray techniques** provide detailed internal scans but process fewer than 20 tubers in 30 minutes, making them impractical for real-time deployment [49,50].
- **RGB CNN classifiers** are lightweight but often behave as black boxes with limited interpretability and single-pass detection only [51,53].
- Our pipeline combines *multi-threshold YOLO*, *ResNet patch verification*, *SAM segmentation*, and *VGG16+RF depth analysis*, reaching F1-scores above 95% with real-time throughput in laboratory conditions using only standard RGB imaging.

This demonstrates that the proposed architecture offers a practical compromise between accuracy and scalability, outperforming traditional RGB deep learning solutions and providing a more accessible alternative to hyperspectral and MRI-based systems.

### 5.5. Discussion

The results confirm that the proposed multi-stage architecture effectively balances high recall with very low false-positive rates, outperforming single-shot CNN classifiers commonly reported in the literature. Unlike hyperspectral or MRI-based techniques, the system operates with standard RGB imaging, making it both cost-effective and industrially scalable. Moreover, the modular design introduces interpretability at each stage, facilitating debugging and adaptation to new defect taxonomies.

Nevertheless, some limitations remain. Rare defects with very few samples (e.g., deep insect galleries) may still challenge the pipeline, suggesting the need for additional data augmentation or synthetic defect generation. Furthermore, the current architecture processes 2D slices only; extending to multi-view or 3D reconstruction could further improve robustness.

Overall, the pipeline demonstrates a strong potential to bridge academic advances in computer vision with the practical requirements of agro-industrial quality control.

## 6. Conclusion

This work presents a hybrid, multi-stage pipeline for the detection and characterization of internal potato defects using cost-effective 2D RGB imaging. By combining high-recall detection with YOLO, patch-level reclassification with ResNet-18, fine-grained segmentation with the Segment Anything Model (SAM), and skin-versus-flesh assessment with a VGG16 + Random Forest module, the system achieves both robustness and interpretability. The results demonstrate that the proposed approach attains high precision and recall while maintaining real-time performance, making it suitable for direct industrial deployment.

Beyond raw performance, the modular design of the pipeline ensures adaptability: each stage contributes complementary strengths that collectively minimize false positives, resolve ambiguous defect classes (such as *mrgal* versus *endo*), and enable accurate quantification of defect extent. This

flexibility provides a framework that can be continuously improved through the integration of new models, datasets, or sensing modalities.

Importantly, the dataset introduced in this study constitutes one of the most diverse and comprehensive collections available for potato quality control, covering multiple varieties including red- and yellow-fleshed cultivars. We anticipate that it will serve as a benchmark reference for future research, enabling rigorous evaluation of AI-based defect detection methods in the agri-food sector.

Future work will explore three directions: (i) extending the approach to other crops such as bananas, (ii) integrating additional imaging modalities (e.g., multispectral or SWIR) for enhanced sensitivity to subsurface defects, and (iii) incorporating uncertainty estimation and explainability modules to further increase industrial trust and adoption.

## References

1. C. Li, Q. Li, G. Chen, Y. Zhang, and Z. Wang, *Identification of Internal Defects in Potatoes Using Hyperspectral Imaging and Convolutional Neural Network*, *Sensors*, vol. 23, no. 9, p. 4065, 2023. doi: [10.3390/s23094065](https://doi.org/10.3390/s23094065)
2. Yan, W., Wang, H., Lu, R., & Guo, W. (2021). Detection of internal defects of potatoes using CNN features and traditional machine learning classifiers. *Computers and Electronics in Agriculture*, 183, 106073. <https://doi.org/10.1016/j.compag.2021.106073>
3. Moallem, P., Dehghani, H., & Omid, M. (2021). Ensemble deep learning model for classification of internal defects in potatoes using visible imaging. *Journal of Food Measurement and Characterization*, 15, 5341–5352. <https://doi.org/10.1007/s11694-021-01036-4>
4. Lee, H. S., & Shin, B. S. (2020). Potato detection and segmentation based on Mask R-CNN. *Computers and Electronics in Agriculture*, 178, 105747. <https://doi.org/10.1016/j.compag.2020.105747>
5. Zhang, Y., Liu, Y., Wang, H., & Lin, J. (2021). Detection of surface defects on potatoes using deep learning and transfer learning techniques. *Agriculture*, 11(9), 863. <https://doi.org/10.3390/agriculture11090863>
6. Qiu, Z., Wang, W., Jin, X., Zhang, H., & Sun, H. (2024). DCS-YOLOv5s: A lightweight algorithm for multi-target recognition of potato seed potatoes based on YOLOv5s. *Agronomy*, 14(11), 2558. <https://doi.org/10.3390/agronomy14112558>
7. Wakholi, C., et al. (2025). HCRP-YOLO: Hybrid Cascade-Refined YOLO for detection of potato sprouting and internal bruises. *Sensors*, accepted manuscript (in press).
8. Alif, M. R. A. R., & Hussain, M. (2024). YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their applications in agriculture. *Artificial Intelligence in Agriculture*, 8, 50–64. <https://doi.org/10.1016/j.aiaa.2024.04.002>
9. Liu, T., Zhou, X., Wang, S., & Yang, R. (2025). YOLOv5s-OURS: A real-time potato surface defect detection model optimized for edge deployment. *Computers and Electronics in Agriculture*, 211, 108174. <https://doi.org/10.1016/j.compag.2024.108174>
10. D. Zhang, J. Liu, J. Sun, Y. Zhao, and Y. Zhang, *Hyperspectral Imaging Combined with Attention-UNet for Internal Defect Detection of Potato Tubers*, *Postharvest Biology and Technology*, vol. 190, p. 111963, 2022. doi: [10.1016/j.postharvbio.2022.111963](https://doi.org/10.1016/j.postharvbio.2022.111963)
11. W. Yan, H. Wang, R. Lu, and W. Guo, *Detection of Internal Defects of Potatoes Using CNN Features and Traditional Machine Learning Classifiers*, *Computers and Electronics in Agriculture*, vol. 183, p. 106073, 2021. doi: [10.1016/j.compag.2021.106073](https://doi.org/10.1016/j.compag.2021.106073)
12. P. Moallem, H. Dehghani, and M. Omid, *Ensemble Deep Learning Model for Classification of Internal Defects in Potatoes Using Visible Imaging*, *Journal of Food Measurement and Characterization*, vol. 15, no. 6, pp. 5341–5352, 2021. doi: [10.1007/s11694-021-01036-4](https://doi.org/10.1007/s11694-021-01036-4)
13. Kamrunnihar, M. and Hashimoto, Y., "Nondestructive Detection of Potato Quality Using LED-Induced Multispectral Imaging," *International Journal of Food Engineering*, vol. 11, no. 4, pp. 551–560, 2015.
14. Rauf, A., Saleem, B. A., and Abbas, Q., "Multispectral Imaging with LED Illumination for Detecting Internal Defects in Potatoes," *International Journal of Food Properties*, vol. 22, no. 1, pp. 1327–1338, 2019.
15. Yang, Y., Liu, Z., Huang, M., Zhu, Q., and Zhao, X., "Automatic Detection of Multi-type Defects on Potatoes Using Multispectral Imaging Combined with a Deep Learning Model," *Journal of Food Engineering*, vol. 336, p. 111213, 2023.
16. Zhang, W., Zhu, Q., Huang, M., and Guo, Y., "Detection and Classification of Potato Defects Using Multispectral Imaging System Based on Single Shot Method," *Food Analytical Methods*, vol. 12, pp. 2920–2929, 2019.

17. Fujikawa, Y., and Kanda, H. (2009). Application of MRI for internal structure investigation in potato tubers. *Journal of Agricultural and Food Chemistry*, 57(12), 5266–5272. <https://doi.org/10.1021/jf900696t>
18. Yan, W., Wang, H., Lu, R., & Guo, W. (2021). Detection of internal defects of potatoes using CNN features and traditional machine learning classifiers. *Computers and Electronics in Agriculture*, 183, 106073.
19. Moallem, P., Dehghani, H., & Omid, M. (2021). Ensemble deep learning model for classification of internal defects in potatoes using visible imaging. *Journal of Food Measurement and Characterization*, 15(6), 5341–5352.
20. Lee, H.-S., & Shin, B.-S. (2020). Potato detection and segmentation based on Mask R-CNN. *Computers and Electronics in Agriculture*.
21. Zhang, et al. (2021). Potato surface defect detection based on deep transfer learning. *Agriculture*, 11(9), 863.
22. Qiu, Z., Wang, W., Jin, X., et al. (2024). DCS-YOLOv5s: A lightweight algorithm for multi-target recognition of potato seed potatoes based on YOLOv5s. *Agronomy*, 14(11), 2558.
23. Wakholi, C., (2025). HCRP-YOLO: A lightweight algorithm for potato defect detection. Preprint.
24. Al Rabbani Alif, M., & Hussain, M. (2024). YOLOv1 to YOLOv10: a comprehensive review of YOLO variants and their application in the agricultural domain. arXiv preprint.
25. Liu, et al. (2025). Improved YOLO v5s-based detection method for external defects in potatoes. *PMCID: PMC11876418*.
26. Schofield, A. M., and Harris, P. J. (2005). Visualizing internal defects in potatoes using X-ray imaging. *Journal of Food Engineering*, 67(3), 301–311. <https://doi.org/10.1016/j.jfoodeng.2004.05.009>
27. Hajjar, G., Quellec, S., Pépin, J., Challos, S., and Musse, M. (2021). MRI investigation of internal defects in potato tubers with particular attention to rust spots induced by water stress. *Postharvest Biology and Technology*, 180, 111600. <https://doi.org/10.1016/j.postharvbio.2021.111600>
28. Collewet, G., Moussaoui, S., Quellec, S., Hajjar, G., and Musse, M. (2023). Characterization of potato tuber tissues using spatialized MRI T<sub>2</sub> relaxometry. *Biomolecules*, 13(2), 286. <https://doi.org/10.3390/biom13020286>
29. Pérez-Torres, E., Kirchgessner, N., Pfeifer, J., and Walter, A. (2015). Assessing potato tuber diel growth by means of X-ray computed tomography. *Plant, Cell & Environment*, 38(11), 2318–2326. <https://doi.org/10.1111/pce.12546>
30. Finney, E. E., and Norris, K. H. (1973). X-ray images of hollow heart potatoes in water. *American Potato Journal*, 50(1), 1–8. <https://doi.org/10.1007/BF02892849>
31. Ibrahim, A., Grassi, M., Lovati, F., Mignani, A. G., and Riminesi, C. (2020). Non-destructive detection of potato tubers internal defects: critical insight on the use of time-resolved spectroscopy. *Advances in Horticultural Science*, 34(1S), 43–51. <https://doi.org/10.13128/ahs-8312>
32. Deng, D., Liu, Z., Lv, P., Sheng, M., Zhang, H., Yang, R., and Shi, T., “Defect Detection in Food Using Multispectral and High-Definition Imaging Combined with a Newly Developed Deep Learning Model,” *Processes*, vol. 11, no. 12, p. 3295, 2023.
33. Semyalo, D., Kim, Y., Omia, E., Arief, M. A. A., Kim, H., Sim, E. Y., Kim, M. S., Baek, I., and Cho, B. K., “Nondestructive Identification of Internal Potato Defects Using Visible and Short-Wavelength Near-Infrared Spectral Analysis,” *Agriculture*, vol. 14, no. 11, p. 2014, 2024.
34. Ruiz, P., García-Vera, V., and Fernández, J., “Challenges in LED-based Multispectral Imaging for Internal Quality Inspection in Tubers,” *Computers and Electronics in Agriculture*, vol. 174, p. 105464, 2020.
35. M. Musse, G. Hajjar, A. Radovic, N. Ali, S. Challos, S. Quellec, P. Leconte, A. Carillo, C. Langrume, L. Bousset-Vaslin, and B. Billiot, *Growth Kinetics, Spatialization and Quality of Potato Tubers Monitored by Magnetic Resonance Imaging*, *Physiologia Plantarum*, vol. 176, no. 3, e14322, 2024. doi: [10.1111/ppl.14322](https://doi.org/10.1111/ppl.14322)
36. FAO, *Potato Market Report 2023*, Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO), 2023. Disponible en ligne : <https://www.fao.org/potato-2023-report> [Consulté le 13 juin 2025].
37. W. Yan, H. Wang, R. Lu, and W. Guo, *Detection of Internal Defects of Potatoes Using CNN Features and Traditional Machine Learning Classifiers*, *Computers and Electronics in Agriculture*, vol. 183, p. 106073, 2021. doi: [10.1016/j.compag.2021.106073](https://doi.org/10.1016/j.compag.2021.106073).
38. Mohd Ali, M., et al. “Quality Inspection of Food and Agricultural Products using Artificial Intelligence.” *\*Reviews in Food Science and Food Safety\**, vol. 18, no. 6, 2021, pp. 1793–1811.
39. Li, C., Li, Q., Chen, G., Zhang, Y., and Wang, Z. (2023). Identification of Internal Defects in Potatoes Using Hyperspectral Imaging and Convolutional Neural Networks. *Sensors*, 23(9), 4065. <https://doi.org/10.3390/s23094065>
40. García-Vera, V., et al. (2024). A Review of Non-Destructive Techniques for Detecting Internal Defects in Agricultural Products. *Postharvest Biology and Technology*, 194, 112024.

41. ElMasry, G., et al. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review. *Food Bioprocess Technol.*, 5, 1121–1142.
42. Kamrunnahar, M., et al. (2015). Internal defect detection in potatoes by LED-based multispectral imaging. *Computers and Electronics in Agriculture*, 116, 64–74.
43. Fujikawa, M., et al. (2009). Evaluation of potato internal defects using magnetic resonance imaging. *Biosystems Engineering*, 102(3), 239–246.
44. Moallem, P., Dehghani, H., and Omid, M. (2021). Ensemble deep learning model for classification of internal defects in potatoes using visible imaging. *Journal of Food Measurement and Characterization*, 15(6), 5341–5352.
45. Yan, W., Wang, H., Lu, R., and Guo, W. (2021). Detection of internal defects of potatoes using CNN features and traditional machine learning classifiers. *Computers and Electronics in Agriculture*, 183, 106073.
46. Arévalo-Royo, J., et al. "AI Algorithms in the Agrifood Industry: Application Potential in the Spanish Agrifood Context. *Applied Sciences*, vol. 15, no.4, 2025, art.2096.
47. D. Zhang, J. Liu, J. Sun, Y. Zhao, and Y. Zhang, *Hyperspectral Imaging Combined with Attention-UNet for Internal Defect Detection of Potato Tubers*, *Postharvest Biology and Technology*, vol. 190, p. 111963, 2022.
48. C. Li, Q. Li, G. Chen, Y. Zhang, and Z. Wang, *Identification of Internal Defects in Potatoes Using Hyperspectral Imaging and CNN*, *Sensors*, vol. 23, no. 9, p. 4065, 2023.
49. Y. Fujikawa and H. Kanda, *Application of MRI for Internal Structure Investigation in Potato Tubers*, *Journal of Agricultural and Food Chemistry*, vol. 57, no. 12, pp. 5266–5272, 2009.
50. G. Hajjar, S. Quellec, J. Pépin, et al., *MRI Investigation of Internal Defects in Potato Tubers with Attention to Rust Spots Induced by Water Stress*, *Postharvest Biology and Technology*, vol. 180, 111600, 2021.
51. W. Yan, H. Wang, R. Lu, and W. Guo, *Detection of Internal Defects of Potatoes Using CNN Features and Machine Learning Classifiers*, *Computers and Electronics in Agriculture*, vol. 183, 106073, 2021.
52. P. Moallem, H. Dehghani, and M. Omid, *Ensemble Deep Learning Model for Classification of Internal Defects in Potatoes Using Visible Imaging*, *Journal of Food Measurement and Characterization*, vol. 15, pp. 5341–5352, 2021.
53. P. Moallem, H. Dehghani, and M. Omid, *Ensemble Deep Learning Model for Classification of Internal Defects in Potatoes Using Visible Imaging*, *Journal of Food Measurement and Characterization*, vol. 15, no. 6, pp. 5341–5352, 2021. doi: [10.1007/s11694-021-01036-4](https://doi.org/10.1007/s11694-021-01036-4).
54. C. Li, Q. Li, G. Chen, Y. Zhang, and Z. Wang, *Identification of Internal Defects in Potatoes Using Hyperspectral Imaging and Convolutional Neural Network*, *Sensors*, vol. 23, no. 9, p. 4065, 2023. doi: [10.3390/s23094065](https://doi.org/10.3390/s23094065).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.