Multi-Scenario Decision Framework for AI Model Selection in Food Visual Inspection

Journal of Food Engineering 422 (2027) 113253

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

Multi-scenario decision framework for AI model selection in food visual inspection

Alberto Piovano [a], Elisa Verna [a,*], Nazarena Cela [b], Luisa Torri [b], Gianfranco Genta [a], Maurizio Galetto [a]

[a] Politecnico di Torino, Department of Management and Production Engineering, Corso Duca degli Abruzzi 24, Torino, 10129, Italy
[b] University of Gastronomic Sciences, Bra, CN, Italy

ARTICLE INFO

Keywords:
Machine vision
Deep learning
YOLO
Food quality control
Cost of quality
Quality 4.0

ABSTRACT

This study addresses the gap between algorithmic performance and manufacturing decision needs in bakery quality control by developing a decision-oriented framework for AI-based visual inspection model selection. Three YOLO object-detection architectures, namely YOLOv5, YOLOv8 and YOLOv11, were benchmarked on a bakery-defect dataset under nominal and controlled synthetic perturbation conditions, with Faster R-CNN included as a non-YOLO baseline. Standard detection metrics, including precision, recall, mAP@50 and mAP@50–95, together with inference latency, were translated into three managerial indicators: Risk-Weighted Total Cost of Quality (TCQ), Throughput Efficiency (TE) and Waste Reduction Index (WRI). These indicators enabled model comparison across high-volume commodity, brand-protection and resource-constrained SME production scenarios. Results show that model suitability is scenario-dependent rather than determined by aggregate detection accuracy alone. YOLOv11 achieved the most favourable profile in high-speed commodity production because its lower latency avoided downtime costs. YOLOv8 provided the lowest TCQ in brand-protection and SME scenarios, where weighted missed-defect costs were more influential than throughput losses. YOLOv5 did not yield the lowest TCQ under the adopted assumptions, but remained relevant as a lower-complexity alternative when false-positive control, waste avoidance or legacy deployment are prioritised. Faster R-CNN was consistently dominated in the investigated case. The proposed cost-quality-throughput decision matrix supports small and medium-sized enterprises (SMEs) in selecting "good-enough" architectures based on constraints and risk appetite rather than technical accuracy alone.

1. Introduction

Industry 4.0 is reshaping food manufacturing, yet quality control still relies heavily on manual inspection. Human evaluation remains essential for subtle sensory defects but is limited by subjectivity, inter-operator variability and fatigue, particularly on high-speed lines (Munoz, 2002). Because key quality variables are often assessed at widely spaced intervals, measurements may not represent the entire batch and defective products can reach consumers between checks (Görgülü, 2025). Ensuring safety, uniformity and sustainability has therefore become a competitive requirement, directly affecting production efficiency, consumer acceptability and brand trust. This is driving the adoption of AI-enabled machine vision (MV) as a Quality 4.0 enabler for real-time quality monitoring and automated inspection. Such systems support the transformation of complex food attributes into objective, data-driven quality metrics (Agrawal et al., 2025, Olakanmi et al., 2023), thereby improving food safety, operational efficiency and sustainability outcomes (Ogidi et al., 2025).

Advances in deep learning (DL) and open-source frameworks have moved MV from laboratory prototypes towards industrial deployment. Convolutional neural networks (CNNs) and the You Only Look Once (YOLO) family learn directly from visual data and can detect complex defects such as baking non-uniformity or texture irregularities with high, task-dependent accuracy in controlled settings (Khan et al., 2020; Olakanmi et al., 2023), supporting real-time monitoring of browning, geometry and hygiene in bakery products (Abdanan Mehdizadeh, 2022; Tantiphanwadi and Malithong, 2022).

Industrial adoption, however, faces a persistent disconnect between algorithmic performance and production economics. Technical studies emphasise incremental gains in mAP or F1-score (Li et al., 2024) but

* Corresponding author.
E-mail address: elisa.verna@polito.it (E. Verna).

https://doi.org/10.1016/j.jfoodeng.2026.113253
Received 13 March 2026; Received in revised form 9 July 2026; Accepted 11 July 2026
Available online 14 July 2026

rarely link detection accuracy to line-level constraints: heavier models may detect more defects yet introduce inference delays that create bottlenecks when inspection time exceeds the line cycle time. Standard metrics also ignore the asymmetric cost of inspection errors in food production, where false positives (FP) mainly generate avoidable scrap, whereas false negatives (FN) may entail external failure, liability and reputational exposure when safety- or compliance-relevant defects are involved (Raak et al., 2017). As a result, manufacturers, particularly small and medium-sized enterprises (SMEs), lack operationally interpretable tools to weigh the cost of latency against the cost of quality risk under explicit production constraints.

This study therefore investigates how AI-based MV model selection can be aligned with manufacturing and food-quality priorities in an industrial bakery inspection case, arguing that architecture choice should be driven by process constraints, error costs and product-quality requirements rather than by aggregate accuracy alone. The paper addresses three research questions:

- RQ1: What are the limitations of current technical metrics in guiding the adoption of deep learning inspection systems in food manufacturing?
- RQ2: How do selected object-detection architectures, including YOLO variants and a non-YOLO baseline, trade off detection performance and inference speed across nominal and controlled synthetic perturbation conditions?
- RQ3: How can these technical performances be translated into managerial indicators integrating cost, throughput and waste-related implications to support evidence-based investment decisions consistent with Quality 4.0 and Zero-Defect Manufacturing (ZDM) strategies?

To address them, a targeted literature review is combined with an experimental case study inspired by bakery conveyor inspection, in which YOLOv5, YOLOv8 and YOLOv11 are benchmarked as representative real-time detectors and Faster R-CNN is included as a non-YOLO baseline. Beyond standard metrics, the analysis applies a decision-oriented Cost-Quality-Throughput framework integrating Risk-Weighted Total Cost of Quality (TCQ), Throughput Efficiency (TE) and Waste Reduction Index (WRI), which operationalises asymmetric inspection risk and the cost of latency to enable context-sensitive comparison of model profiles. The purpose is not to identify a single best-performing architecture across all scenarios, but to provide a recalibrable decision logic for aligning model selection with product characteristics, line constraints and cost-of-quality priorities.

2. Literature review

2.1. Corpus selection

The literature review is based on a PRISMA-guided search and screening protocol (see Fig. 1). The resulting final corpus (n = 18) is analysed through three complementary lenses: (i) image acquisition and pre-processing (Table 1), (ii) algorithmic families and deployment trade-offs (Table 2), and (iii) defect typologies and quality attributes (Table 3). Therefore, Tables 1–3 refer to the same set of included studies, viewed from different perspectives. The predominance of bakery and

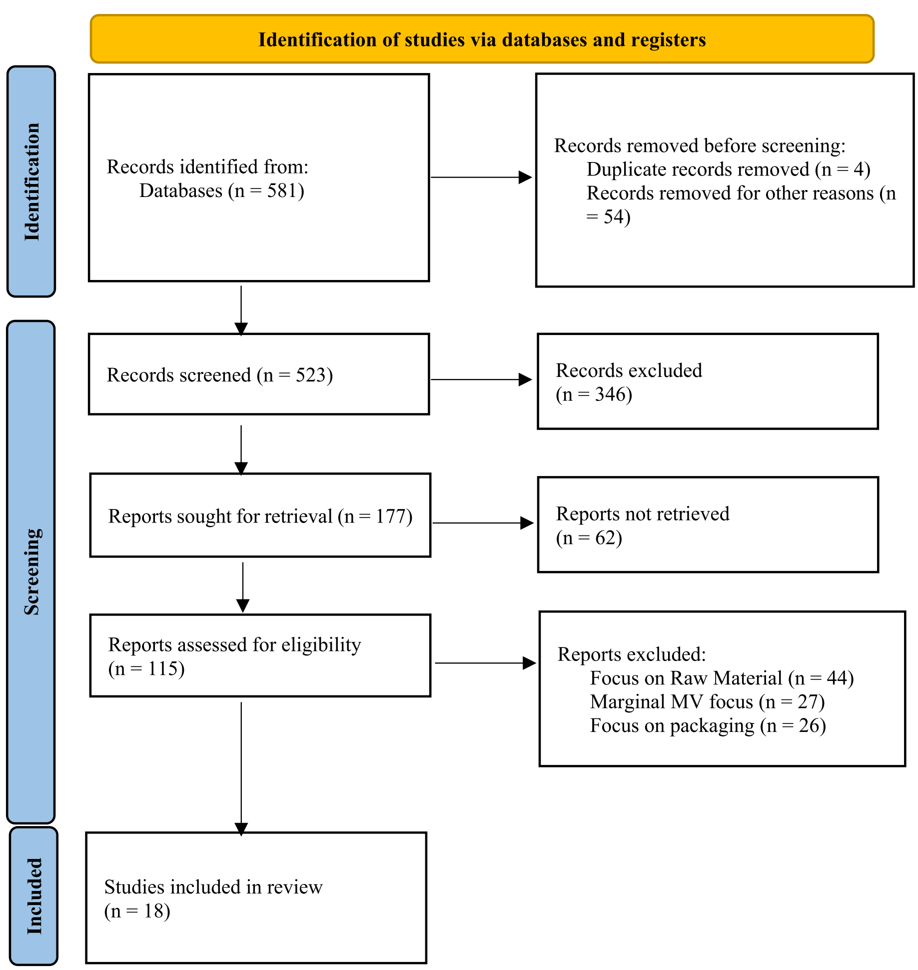


Fig. 1. PRISMA Flow Diagram. This diagram illustrates the study selection process for the literature review.

Table 1
Image acquisition and pre-processing techniques in food manufacturing quality inspection applications.

Acquisition Type	Image Resolution [pixels]	Experimental Conditions	Application Context	Algorithm/ Processing	Ref.(s)
Digital camera	256x256, 150x150	Camera at 90° to object; two D65 light sources at 45°. Vertical setup at 24 cm with dual flashlights.	Conveyor-belt image capture for real-time inspection	Machine vision, CNN, machine learning	(Abdanan Mehdizadeh, 2022; Drogalis et al., 2024; Gökmen and Mogol, 2010; Mogol and Gökmen, 2014; Rezagholi and Hesarinejad, 2017; Rohan Joel and Benjamin, 2024; Tantiphanwadi and Malithong, 2022)
Optical camera	1024x768	Natural lighting, no artificial source.	Single-image geometric feature extraction	Support vector machine (SVM)	Mannaro et al. (2022)
Overhead camera	227x227x3 2000x2000 2376x584	Camera at 90° to objects under controlled illumination.	Conveyor-belt acquisition for surface analysis	CNN	(Kılcı and Koklu, 2025; Li et al., 2024; Tutuncu et al., 2023)
Blackout chamber	1024x512	Vertical camera orientation; LED illumination for uniform light field.	Single-image capture during baking process	Machine learning	Lee et al. (2023)

Table 2
Overview of traditional machine learning (feature-based) methods, deep learning (CNN-based) models and YOLO-family object detectors, highlighting typical tasks, computational load, main advantages and limitations from an industrial deployment perspective.

Category/ Model	Technology	Task/Application	Reported computational requirement	Advantages	Limitations	Reference(s)
Deep learning (CNN)	VGG16 (Visual Geometry Group 16-layer network)	Transfer learning and fine-tuning with image augmentation for feature extraction and classification.	Training time: 42.1 min (30 epochs), single GPU.	- Established and reliable architecture. - Suitable for classification and pre-processing tasks.	- High memory and training time requirements. - Increases operational and energy costs.	(Tantiphanwadi and Malithong, 2022; Tutuncu et al., 2023)
Deep learning (YOLO object detectors)	YOLOv3	Object detection under edge computing conditions.	GPU: NVIDIA RTX 2080 Super, 64 GB RAM.	-Real-time detection and good inference speed. -Suitable for embedded and low-latency applications.	-Limited accuracy for small or overlapping objects. -Degraded performance in visually dense environments.	Chen et al. (2022)
	YOLOv5	Object detection and foreign body identification.	NA	-Balanced trade-off between speed and accuracy. -Broad community support and documentation. -Compatible with mainstream GPU.	-Suboptimal on complex datasets. -Limited scalability for high-variability industrial environments.	Rohan Joel and Benjamin (2024)
	YOLOv7	Object detection and aesthetic quality inspection.	10.9 ms per frame; 73.1 MB model size.	-High GPU efficiency. -Improved accuracy. -Suitable for automated production lines.	-Increased architectural complexity -Requires dedicated hardware and expert configuration.	Li et al. (2024)
	YOLOv8	Object detection and cooking stage recognition (browning stage classification).	NA	-Unified framework for detection, segmentation, and classification. -High flexibility and user-friendly interface.	-Requires advanced hyperparameter tuning. -Dependent on well-annotated datasets and robust validation pipelines.	Kılcı and Koklu (2025)
Traditional Machine Learning	SVM	Image segmentation and feature-based classification.	Evaluation period: Jan 2020–Mar 2021.	-High accuracy on small datasets. -Effective in binary problems.	-Limited scalability. -Sensitive to kernel and parameter selection. -Manual tuning for high-dimensional datasets.	Mannaro et al. (2022)
	Random Forest (RF), k-Nearest Neighbors (KNN)	Panel volume, colour measurement and statistical analysis.	Arduino Uno hardware integration	-Direct integration with industrial machinery. -Effective for process monitoring.	-Sensitive to environment. -Requires frequent calibration to ensure reproducibility.	(Khan et al., 2020; Lee et al., 2023)

confectionery studies reflects both the distribution of the retrieved industrial MV evidence and the relevance of these categories to the downstream empirical case.

2.2. Machine vision for quality assurance

In industrial contexts, MV can be understood as the operational extension of computer vision to inspection systems embedded in production lines (Terven et al., 2023). By combining image acquisition, controlled illumination, processing algorithms and, where required, sorting or rejection devices, MV enables non-destructive assessment of shape, colour, surface texture and contamination, providing a scalable complement to manual inspection, which can be labour-intensive, subjective and error-prone.

Food processing is an important domain for MV because product safety, uniformity and efficiency are increasingly critical competitive

Table 3
Mapping recurrent defect types (cooking, hygienic/compliance, geometric and overall aesthetic quality) to representative products, algorithms, computer-vision tasks, performance metrics and dataset sources.

Defect Type	Description	Products Analysed	Algorithms Used	Vision Task	Performance Metrics	Dataset	Reference(s)
Cooking Defects	Assessment of cooking stage (undercooked/optimal/overcooked) based on thermal and/or colour threshold.	Cupcakes, biscuits, bread, cake, fried potatoes	CNN (VGG16), ML	Image segmentation, transfer learning, colour and volume measurement	Browning ratio, accuracy, precision, recall, F1-score, p-value, volume	Private and public	(Gökmen and Mogol, 2010; Mogol and Gökmen, 2014; Abdanan Mehdizadeh, 2022; Tantiphanwadi and Malithong, 2022)
Hygienic/Compliance Defects	Detection of gluten presence and potential mould contamination.	Cake, bread	MV, CNN (YOLOv5)	Crumb and crust image processing; defect identification and classification	L*, a*, b* colour parameters, precision, recall, F1-score	Private dataset	(Rezagholi and Hesarinejad, 2017; Rohan Joel and Benjamin, 2024)
Geometric Defects	Surface imperfections (bubbles, cracks, stains), shape irregularities, and incorrect ingredient distribution.	Candy, biscuits, *pane carasau*	CNN (YOLOv3/v7/v8), SVM, ML	Segmentation and classification; defect identification; feature extraction	Precision, recall, F1-score, mAP@50, mAP@50–95, confusion matrix	Private and public datasets	(Chen et al., 2022; Mannaro et al., 2022; Drogalis et al., 2024; Li et al., 2024; Kılcı and Koklu, 2025)
Overall Aesthetic Quality	Evaluation of product conformity and marketability (sellability, visual appeal).	Milk powder, bread, *taralli* biscuits	MV, ML	Global classification of defect classes; aesthetic scoring	Precision, recall, F1-score, mAP@50, mAP@50–95, confusion matrix	Private dataset	(Khan et al., 2020; Tutuncu et al., 2023)

levers (Piovano et al., 2026). As consumer awareness of quality and sustainability intensifies, manufacturers must guarantee sensory consistency and hygienic quality while controlling costs and reducing waste (Xiao et al., 2022; Feng et al., 2025). MV systems address this need through in-line inspection on conveyors and along thermal-processing stages, supporting earlier defect detection, rework reduction and feedback to process control. Applications range from monitoring surface browning during baking, associated with Maillard-driven colour development and the underlying heat- and mass-transfer processes governing crust formation (Purlis, 2010), to detecting foreign bodies and hygiene-related non-conformities, to assessing geometric and aesthetic conformity of high-volume bakery products (Abdanan Mehdizadeh, 2022; Li et al., 2024; Srivastava et al., 2014; Tutuncu et al., 2023). Representative acquisition and pre-processing configurations are summarised in Table 1, which is illustrative rather than exhaustive.

2.3. From traditional machine vision to YOLO

Early MV applications in food relied on handcrafted features and traditional classifiers such as Support Vector Machines (SVM), Random Forests (RF) or k-Nearest Neighbors (KNN) to perform tasks like colour grading, segmentation and basic defect classification in bread, biscuits and other baked products. These approaches can perform well on small, well-controlled datasets and are compatible with low-cost hardware, but they scale poorly to high-variability, real-time industrial environments and require substantial parameter tuning (Reichenstein et al., 2022). From a Quality 4.0 perspective, such rule-based systems reflect early stages of digitalisation, in which quality data remain fragmented and only partially integrated with higher-level information systems, limiting automated, evidence-based decision-making across the value chain (Chiarini & Kumar, 2022).

Deep learning has progressively replaced these pipelines with end-to-end architectures. CNNs have been successfully used for classifying baking stages and aesthetic quality (Tantiphanwadi and Malithong, 2022; Tutuncu et al., 2023). However, many industrial food-inspection tasks require not only image-level classification, but also real-time localisation of multiple defects on moving products, which makes end-to-end detection architectures more directly applicable in conveyor-based settings.

More recently, the YOLO family of object detectors has become prominent in real-time food inspection because it often offers a favourable balance between detection capability and inference speed. As summarised in Table 2, recent studies frequently report competitive performance of YOLO variants in visually complex inspection tasks (J. Chen et al., 2022; Kılcı and Koklu, 2025; Li et al., 2024; Rohan Joel and Benjamin, 2024). At the same time, the move towards end-to-end learning and automated feature extraction is aligned with Quality 4.0's emphasis on automatic data collection and vertical/horizontal integration of quality information with enterprise resource planning (ERP), manufacturing execution systems (MES) and supervisory control and data acquisition (SCADA) systems for real-time, evidence-based quality control (Chiarini & Kumar, 2022).

These performance gains come with higher demands in terms of GPU availability, dataset quality and specialised skills for training and tuning. As summarised in Table 2, the literature portrays a clear trade-off: lightweight, traditional algorithms minimise upfront investment but offer limited robustness and generalisability, whereas DL approaches, and YOLO-based detectors in particular, capture complex visual non-conformities more effectively but require greater data, hardware and organisational resources. This resonates with Quality 4.0 adoption studies, which identify perceived cost and time savings and the availability of enabling infrastructures as key drivers for investing in digital quality tools, especially in resource-constrained settings (Sony and Naik, 2020; Talaie et al., 2024). What remains underexplored is how the accuracy-latency trade-off changes when detectors are evaluated under explicit resource and deployment constraints relevant to industrial food inspection.

2.4. Defect typologies and quality attributes

After considering the main algorithmic families, it is necessary to clarify which defects MV systems are expected to detect and how these defects relate to economic and quality outcomes. While the review covers food MV broadly, this subsection emphasises bakery and confectionery because these domains are both highly represented in the screened corpus and directly aligned with the empirical case study developed in Section 4. In bakery and confectionery applications, studies converge on a relatively stable set of defect typologies. Using the same screened corpus underlying Tables 1 and 2, Table 3 synthesises these recurrent defects, linking them to representative products, algorithms and performance metrics. Although the case study discussed in Section 4 is drawn from a bakery context, the defect families synthesised in Table 3 are not product-specific. Thermal and colour deviations, structural or geometric defects, compliance defects, and overall visual conformity recur across multiple food categories, including baked goods, confectionery, and ready-to-eat foods. Therefore, any transferability of the proposed framework should be considered primarily at the level of recurrent defect families, inspection objectives, and risk-bearing quality categories, rather than assumed at the level of a specific product or process context.

Four main clusters can be distinguished:

1. Cooking defects, referring to undercooked, optimal and overcooked products identified through crust and crumb colour and texture. These markers are closely associated with Maillard reaction progress and heat load, influencing flavour development and textural changes in the crust. Therefore, MV-based browning assessment can support more consistent baking and thermal processing, reducing rework and scrap while supporting process monitoring relevant to acrylamide-mitigation strategies (Capuano et al., 2008; Abdanan Mehdizadeh, 2022; Lee et al., 2023).
2. Hygienic and compliance defects, including visible contamination, mould-related anomalies, allergen-related non-conformities and other compliance-relevant defects in bread, cakes and dairy, where MV may complement laboratory testing by enabling earlier in-line screening and rejection (Lee et al., 2023; Tutuncu et al., 2023). The non-destructive detection of these defects is vital not only for ensuring food safety but also for protecting company reputation, as foreign contaminants pose risks to both consumer health and brand image (Yin et al., 2021).
3. Geometric defects, such as irregular shapes, bubbles, cracks, stains or non-uniform ingredient distribution in biscuits, candies and traditional breads. These attributes directly influence visual appearance, which represents the dominant and often primary sensory modality in human perception, useful for initial product evaluation (Hutmacher, 2019). YOLO-based detectors and SVMs are frequently adopted to classify and localise such defects on high-speed conveyors (Li et al., 2024; Mannaro et al., 2022).
4. Overall aesthetic quality, defined as global "sellability" or visual conformity, typically encoded as multi-class quality labels (e.g. acceptable, rework, reject) and directly linked to customer perception and product positioning.

These defect typologies are not merely technical categories; they act as proxies for scrap, rework, food-safety risk and brand perception. In cost-of-quality terms, cooking and geometric defects are more commonly associated with internal failure costs, whereas hygienic or compliance-related defects may generate more severe external failure consequences, including recalls and reputational damage (Verna et al., 2022).

Technical performance is usually reported through precision, recall, F1-score, confusion matrices and, for object detection, mAP@50 or mAP@50–95 (Everingham et al., 2010; Senni et al., 2014). Some studies also incorporate physical attributes (e.g. volume, aspect ratio) for baked goods. In parallel, economic inspection models connect these error rates to inspection, rework and failure costs, and propose composite indicators such as the Return on Investment of Inspections (ROII) to evaluate inspection strategies and support ZDM goals (Verna et al., 2022).

2.5. Adoption and the managerial gap

Beyond academic prototyping, MV is increasingly deployed in industrial lines as a Quality 4.0 enabler, yet a persistent gap remains between the technological maturity of algorithms and the frameworks guiding their adoption. This study addresses three such gaps: (i) the limitations of technical metrics alone for ZDM-oriented decisions and the lack of risk- and latency-aware evaluation (Section 2.5.1); (ii) the scarcity of comparative, decision-oriented evidence on "good-enough" architectures under SME constraints (Section 2.5.2); and (iii) the fragmented integration of Quality 4.0 principles with concrete inspection design choices (Section 2.5.3).

2.5.1. Limits of technical metrics in ZDM

Most computer-vision studies still evaluate models almost exclusively through technical indicators. These metrics are necessary for engineering validation but are insufficient for ZDM-oriented decision-making. ZDM research stresses that defect elimination must be balanced against resource consumption and cost-of-quality trade-offs, rather than pursued as an abstract "zero defect" ideal (Psarommatis et al., 2020). Recent cost-modelling work on inspection reinforces this view by explicitly integrating inspection costs, rework and replacement costs, and the costs of poor quality due to inspection errors into a single performance measure, thus showing that Type I and Type II errors have structurally different economic impacts (Verna et al., 2022).

In food inspection, missing a safety- or compliance-relevant defect (FN) can generate external failure and reputational losses that greatly exceed the cost of scrapping borderline products (FP). In premium or brand-sensitive contexts, even major sensory non-conformities may carry non-negligible commercial consequences, although typically with a different severity profile. Moreover, the emerging Quality 4.0 literature defines flawless, highly automated procedures as a key objective (Radziwill, 2020; Talaie et al., 2024), but it rarely quantifies the "Cost of Latency": the economic penalty incurred when a technically accurate model is too slow to keep up with line cycle times, forcing buffers, rework loops or reduced throughput. Studies on AI-enabled in-process inspection in other manufacturing contexts highlight that hardware acceleration may be necessary to prevent the inspection system itself from becoming a production bottleneck (Villalba-Diez et al., 2019). Overall, managers are left without operational tools to translate technical performance (mAP, latency) into risk-weighted cost-of-quality indicators that are coherent with ZDM strategies.

2.5.2. Limited comparative evidence on "good-enough" architectures for SMEs

A second gap exists in the level of analysis applied to digital quality technologies. Existing research tends to be either broadly conceptual (focusing on maturity models or systemic enablers) or narrowly technical, optimizing a single algorithm on a proprietary dataset (Sony and Naik, 2020; Chiarini & Kumar, 2022). This leaves a methodological "missing middle" where concrete, practical decisions between competing technologies are seldom examined. For SME food manufacturers, however, the critical challenges are more specific: deploying systems on limited-capacity edge devices, operating in harsh environments, and relying on multi-skilled staff instead of data science specialists. Consequently, the central question for these firms shifts from achieving peak benchmark accuracy to identifying a model that provides "sufficient" performance within strict constraints of cost, energy use, and skill requirements. While economic models (e.g., ROII frameworks) exist to rank inspection strategies by cost-of-quality, they are rarely integrated with systematic, operationally interpretable comparisons of deep-learning architectures in food-production settings. This disconnect leaves an important evidence gap for selecting viable, "good-enough" technologies aligned with the capabilities and resource profiles of SMEs.

2.5.3. Fragmented integration of quality 4.0

Finally, while Quality 4.0 research rightly stresses that digital tools must align with strategy, culture, and customer objectives to create value, a critical disconnect persists. Existing frameworks highlight integrated data flows and automated processes to enable evidence-based decisions (Chiarini & Kumar, 2022; Talaie et al., 2024). Yet, few studies explicitly bridge these strategic imperatives with the concrete technical choices involved in designing an automated inspection system, such as selecting an inspection mode (offline, in-line, mixed-reality) or a specific model architecture. Evidence indicates that perceived costs, infrastructure, and organisational resistance shape investment decisions, especially under constraints. However, there is little guidance on how different technical configurations (e.g., a lightweight vs. a high-accuracy model) support fundamentally different manufacturing strategies, such as high-throughput commodity production versus premium, brand-protective operations. This fragmentation risks preventing machine vision from evolving into a coherent strategic capability, relegating it instead to the status of an isolated technical project.

2.5.4. Contribution of this study

Addressing these gaps, this study contributes a decision-oriented framework for translating AI-based visual inspection performance to manufacturing strategy and cost-of-quality implications. Building on ZDM-oriented cost modelling and Quality 4.0 frameworks, it combines a PRISMA-guided synthesis of food MV evidence with an experimental bakery inspection case benchmarking multiple YOLO variants, plus a non-YOLO baseline, under nominal and controlled synthetic perturbation conditions. The emphasis is placed on translating model performance into operationally interpretable indicators, rather than on proposing a new detection architecture. By jointly evaluating technical indicators (precision, recall, mAP, latency) and managerial indicators (Risk-Weighted TCQ, TE and WRI), it links asymmetric inspection risk, throughput feasibility and waste-related implications within a single cost-quality-throughput logic. TCQ, TE and WRI are not introduced as new theories, but as adapted decision-support indicators that translate model performance into manufacturing relevance under explicit assumptions (Fig. 2). The framework supports structured comparison among model profiles across production priorities and resource constraints, while acknowledging that transferability beyond the investigated bakery case requires further validation.

3. Methodology

To bridge the gap between algorithmic performance and manufacturing strategy, this study adopts a multi-stage research design comprising model selection and experimental setup (Sections 3.1-3.2), a hierarchical performance evaluation framework (Section 3.3) and an economic scenario analysis (Section 3.4).

3.1. Research design

The selection of the YOLO framework is motivated by process compatibility with real-time industrial inspection rather than by algorithmic novelty. Three YOLO generations were selected to sample the accuracy-latency trade-off within a real-time, single-stage detection family: YOLOv5, as a mature and widely adopted model; YOLOv8, as a more recent anchor-free architecture; and YOLOv11, as a newer iteration designed for improved computational efficiency (Cai et al., 2025). The full set of technological and organisational criteria that guided model inclusion is summarised in Table 4.

The benchmark was centred on these three YOLO variants because the objective was to compare alternative real-time profiles under operational constraints, not to exhaustively review the state of the art in object detection. However, to avoid limiting the empirical evidence to a single detector family, Faster R-CNN was added as a conservative non-YOLO baseline (Ren et al., 2017). Its inclusion provides a useful reference point for assessing whether a two-stage detector, typically associated with stronger localisation logic but higher computational burden, would offer a competitive cost-quality-throughput profile in the investigated setting. The four models should therefore be interpreted as a structured comparison between three real-time YOLO alternatives and one established non-YOLO baseline, rather than as an exhaustive model-selection study.

Table 4
Technological and organisational criteria guiding the inclusion of YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN in the empirical study, with a focus on real-time capability, architectural comparability, and relevance for small and medium-sized food manufacturers.

Selection Criterion	Description and SME-oriented Rationale	Models Included/ Excluded
Framework support and stability	SMEs require reliable, low-maintenance repositories.	*Included:* YOLOv5, YOLOv8, YOLOv11. *Excluded:* poorly maintained or obsolete YOLO releases.
Representation of architectural evolution	Multiple YOLO generations were included to capture changes in the accuracy-latency trade-off within a comparable real-time detection family.	*Included*: YOLOv5, YOLOv8, YOLOv11.
Accuracy–latency trade-off	Real-time inspection requires detection performance to be evaluated jointly with inference speed, since slow models may become production bottlenecks.	*Included:* all YOLO variants and Faster R-CNN.
Non-YOLO reference baseline	A two-stage detector was included to provide a conservative architectural benchmark beyond the YOLO family.	*Included*: Faster R-CNN.

3.2. Experimental setup and synthetic perturbation protocol

All experiments were conducted on an NVIDIA T4 GPU. This computational setting was selected to provide a constrained and operationally relevant environment, rather than to emulate a specific industrial edge device. The YOLO models were implemented using Ultralytics workflow, while Faster R-CNN was included as a non-YOLO

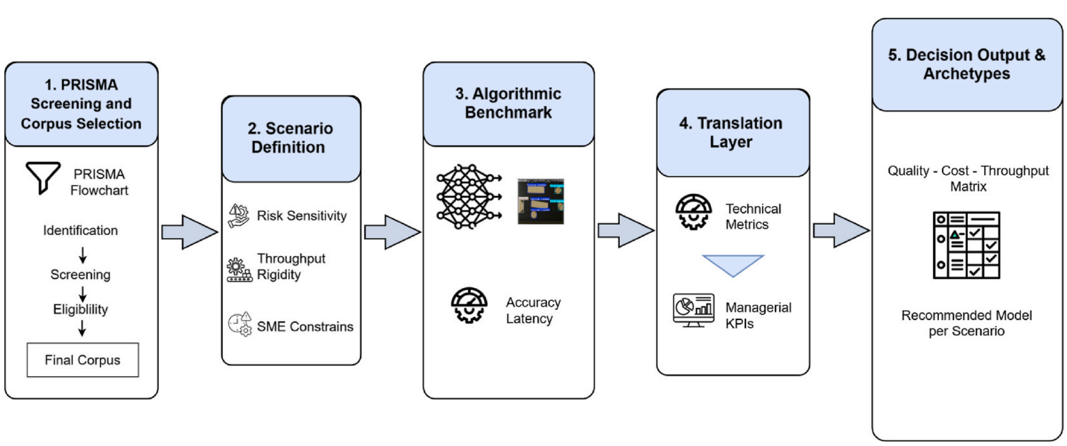


Fig. 2. Schematic overview of the multi-stage research methodological framework.

baseline and evaluated under the same dataset conditions. To reduce the dependence of the results on a single random initialisation, each model-condition combination was evaluated over three independent runs. The main YOLO training settings are reported in Appendix Table A1, while the corresponding YOLO training and validation loss curves are provided in Appendix Figure A1 as convergence diagnostics. Faster R-CNN was implemented in PyTorch and evaluated under the same dataset conditions. Training was aligned with the YOLO experiments in terms of epochs and batch size, while inference latency was measured per image on the same NVIDIA T4 GPU.

The empirical study uses a publicly available bakery-defect dataset composed of cracker images acquired in a conveyor-belt inspection setting. The dataset was selected for its relevance to conveyor-based visual inspection and its compatibility with object-detection annotation. The original release included two labelled classes, Normal and Cracked. For the purposes of the present study, all images and annotations were manually reviewed, the existing labels were checked, and a third class, Over-baked, was manually added to enable a three-class defect analysis aligned with the study objectives. In the present work, the Over-baked class is treated as a visual proxy for severe thermal-processing non-conformities associated with excessive browning and potential risk-bearing quality deviations in bakery products (Purlis, 2010). To compare model behaviour under less ideal visual conditions, a controlled synthetic perturbation protocol was implemented (Table 5). The protocol applies shear and Gaussian blur to generate a synthetically perturbed dataset condition. These perturbations are used as a reproducible methodological proxy for selected acquisition degradations consistent with augmentation-based evaluation practices in applied deep learning (Mumuni and Mumuni, 2022), rather than as a physical simulation of specific industrial disturbances.

The two dataset conditions were defined as follows. The Nominal condition used standard preprocessing, including auto-orientation and resizing to 640 × 640 pixels, and served as the reference condition. The Synthetic Perturbation condition used the same base corpus reprocessed with a more aggressive augmentation schedule, including horizontal flip, shear and blur. This condition was designed to increase visual variability and partially reduce class asymmetry. It should therefore be interpreted as a controlled comparison between two dataset conditions, not as a pure test-time robustness benchmark.

Class imbalance was addressed through manual relabelling, dataset restructuring and controlled augmentation, with the aim of increasing variability and partially reducing class asymmetry rather than implementing advanced reweighting or resampling strategies. Given that an independent external validation dataset was not included and that the test subsets were relatively modest in size, the findings should be interpreted as evidence from a controlled three-class bakery case study. The objective is not to provide a general benchmark across food categories, but to examine how technically comparable detectors, complemented by a non-YOLO baseline, lead to different managerial implications once error type, throughput constraints and scenario-specific cost assumptions are considered.

3.3. Performance evaluation framework

The evaluation follows a hierarchical structure designed to link technical metrics directly to managerial outcomes.

3.3.1. Technical metrics

Technical performance is characterised by Precision, Recall, mAP@50 and mAP@50-95, and end-to-end latency (ms), following established computer vision evaluation protocols (Terven et al., 2023; Dikici and Robinson, 2024). Consistent with the repeated-run protocol described in Section 3.2, all aggregate metrics are reported as mean ± standard deviation over three independent training runs (Table 7), while class-level metrics for each model and condition are provided in Appendix Table A2.

3.3.2. Managerial key performance indicators (KPIs)

To translate technical performance into organisational value, three specific KPIs are introduced, building on the approaches of Wuest et al. (2016). The three KPIs are presented as complementary decision-support indicators addressing three distinct managerial questions: the economic burden of inspection errors (Risk-Weighted TCQ),

Table 6
Stylised operational assumptions and scenario-dependent decision weights used in the comparative economic simulation (Commodity, Brand Protection, SME).

Parameter	Measurement Unit	Scenario A: Commodity (High Volume)	Scenario B: Brand Protection (Quality Focus)	Scenario C: SME (Resource Constrained)
Strategic Priority	-	Speed & Throughput	Zero-Defect/ Compliance	Cost & Waste Reduction
Batch Size	Units	100000	50000	20000
Target Line Speed	Frames per second (FPS)	50 (High Speed)	30 (Moderate)	45 (Standard)
C_{scrap}	€/unit	0.01	0.05	0.02
C_Q	€/unit	0.10 (Low Impact)	1.00 (Brand Impact)	0.25 (Local Reputation)
C_S	€/unit	1.00 (Regulated)	5.00 (Critical)	2.00 (High)
$C_{downtime}$	€/hour	1000 (Major Loss)	500 (Manageable)	200 (Low)
Operational Rationale	-	High-saturation lines where latency causes direct revenue loss via T_{loss}	Premium segments where safety failures $N_{FN,S}$ trigger severe penalties.	Smaller batches where hardware budget is limited and waste disposal (FP) is a key KPI.

Note: Cost parameters are scenario-dependent decision weights. $C_{downtime}$ denotes the marginal opportunity cost of latency-induced production-time loss. Their hierarchy $C_{scrap} < C_Q < C_S$ follows CoQ/PAF logic, distinguishing internal scrap from aesthetic external failures and safety-related external failures.

Table 5
Composition and augmentation characteristics of the two dataset conditions used for the comparative experiments.

Data Property	Nominal Set	Perturbation Set	Rationale
Total images	658	1045	Higher variability for controlled robustness testing and partial reduction of class asymmetry.
Split (train/val/ test)	537/81/40	924/81/40	Comparable validation/test sets for aligned evaluation across the two dataset conditions.
Pre-processing	Auto-orient; resize 640×640	Auto-orient; resize 640×640	Uniform image scale.
Data augmentation	Horizontal flip	Horizontal flip; shear ±10° H/V; blur 4.3 px	Introduces controlled synthetic distortions and blur compatible with conveyor-based inspection scenarios.

Note: Object-level class distributions for each dataset split are reported in Appendix Table A3. Because the task is object detection rather than image-level classification, multiple product instances may be present within a single frame.

Table 7
Technical performance benchmarking of YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN under Nominal and Synthetic Perturbation conditions. Values are reported as mean ± standard deviation over three independent training and evaluation runs initialized with different random seeds.

Model	Scenario	Precision (P)	Recall (R)	mAP@50	mAP@50-95	Total Latency [ms]
YOLOv5	Nominal	0.915 ± 0.015	0.945 ± 0.028	0.971 ± 0.008	0.951 ± 0.004	27.6 ± 0.7
	Perturbation	0.909 ± 0.054	0.921 ± 0.027	0.971 ± 0.007	0.947 ± 0.004	27.4 ± 1.0
YOLOv8	Nominal	0.904 ± 0.013	0.913 ± 0.012	0.963 ± 0.012	0.941 ± 0.009	26.2 ± 1.9
	Perturbation	0.889 ± 0.036	0.940 ± 0.050	0.971 ± 0.009	0.945 ± 0.007	24.8 ± 0.8
YOLOv11	Nominal	0.896 ± 0.008	0.940 ± 0.024	0.966 ± 0.004	0.945 ± 0.008	18.4 ± 0.6
	Perturbation	0.931 ± 0.020	0.915 ± 0.021	0.976 ± 0.002	0.949 ± 0.002	14.6 ± 0.1
Faster R-CNN	Nominal	0.820 ± 0.022	0.895 ± 0.036	0.869 ± 0.030	0.747 ± 0.035	28.7 ± 2.4
	Perturbation	0.762 ± 0.036	0.922 ± 0.012	0.868 ± 0.014	0.701 ± 0.027	27.6 ± 0.8

the ability to sustain the required production rate (TE), and the prevention of unnecessary waste (WRI).

(i) Economic Impact: Risk-Weighted Total Cost of Quality (TCQ)

Adapting traditional Cost of Quality (CoQ) models, and specifically the Prevention-Appraisal-Failure (PAF) logic that distinguishes internal from external failure costs, a Risk-Weighted TCQ is proposed (Farooq et al., 2017; Plunkett and Dale, 1987; Porter and Rayner, 1992). This model decomposes external failure costs into distinct tiers based on defect severity, a critical distinction in food manufacturing, where missing a safety hazard has fundamentally different financial consequences than missing an aesthetic flaw. The TCQ per batch is calculated as:

$$TCQ = \left(N_{FP} \cdot C_{scrap}\right) + \left(N_{FN,Q} \cdot C_Q\right) + \left(N_{FN,S} \cdot C_S\right) + \left(T_{loss} \cdot C_{downtime}\right) \tag{1}$$

Where:

- N_{FP}: Number of compliant items incorrectly rejected (Internal Failure) [units].
- C_{scrap}: Unit production cost (materials, energy) lost due to scrap [€/units].
- $N_{FN,Q}$ (Aesthetic Risk): Number of "Cracked" items missed. These represent functional/aesthetic defects affecting consumer satisfaction but not safety [units].
- C_Q: Cost of non-conformance for aesthetic defects (e.g., customer complaints, minor refunds) [€/units].
- $N_{FN,S}$ (Safety Risk): Number of "Over-baked" items missed. These act as a proxy for chemical contamination risks (see Section 3.3) [units].
- C_S: Cost of non-conformance for safety defects (e.g., recalls, regulatory penalties, brand damage) [€/units].
- T_{loss} (Downtime Loss): Production time lost if the model's inference speed is slower than the line target [h].
- $C_{downtime}$: Hourly opportunity cost of production stoppage [€/h].

It is worth noting that the "Cost of Latency" ($T_{loss} \cdot C_{downtime}$) explicitly monetizes production losses caused by algorithms slower than the line target, a variable often neglected in computer vision literature (Hanhirova et al., 2018). In practical terms, TCQ increases when the model misses costly defects or generates unnecessary rejects. Lower values indicate better risk-cost performance.

(ii) Operational Impact: Throughput Efficiency (TE)

Adapted from the Overall Equipment Effectiveness (OEE) tradition (S.Nakajima, 1988), TE measures the system's ability to maintain the target line speed without buffering:

$$TE = min\left(1, \frac{FPS_{model}}{FPS_{target}}\right) \tag{2}$$

where FPS indicate frames per second.

A $TE < 1$ indicates that the inspection system has become the process bottleneck (e.g., at 40 FPS with a 50 FPS target, TE = 0.80).

(iii) Environmental Impact: Waste Reduction Index (WRI)

Aligned with Circular Economy logic, WRI quantifies the relative reduction of avoidable scrap associated with FP.

$$WRI = \frac{FP_{baseline} - FP_{model}}{FP_{baseline}} \tag{3}$$

Where $FP_{baseline}$ is the number of FP count of the reference configuration adopted for comparison, while FP_{model} is the number of FP count of the evaluated model. This index directly links algorithmic precision to tangible sustainability outcomes (Lehn and Schmidt, 2022). Higher values indicate better prevention of avoidable waste.

3.4. Economic modelling and scenarios

Cost parameters were specified according to the differing severity of inspection errors under alternative production priorities. Accordingly, the cost assigned to a missed risk-bearing defect is assumed to reflect not only unit product loss, but also potential compliance actions, recall exposure and reputational damage, whose economic impact may exceed product value by orders of magnitude (Gomez and Marks, 2020).

Within the CoQ and PAF logic, cost parameters are specified as scenario-dependent decision weights rather than audited plant-level accounting values. Their purpose is to preserve plausible order-of-magnitude differences between avoidable scrap, aesthetic external failures and risk-bearing external failures, enabling comparative model ranking under explicit strategic assumptions rather than exact plant-cost estimation (Muffato Reis et al., 2025; Plunkett and Dale, 1987; Psomas et al., 2018). The downtime parameter $C_{downtime}$ is interpreted as the marginal opportunity cost of latency-induced production-time loss, consistent with bakery scheduling studies showing that makespan and oven idle time materially affect manufacturing costs, and with continuous-baking models in which conveyor speed determines production rate (Babor et al., 2023; Broyart and Trystram, 2002).

Three stylised manufacturing scenarios were defined to compare model suitability under different strategic priorities rather than to reproduce a specific plant configuration (see Table 6). Scenario A represents high-volume commodity production, where the priority is maintaining throughput and avoiding downtime on a 50 FPS line. Scenario B represents brand-protection or premium production, where the priority is reducing the escape of risk-bearing defects under moderate line-speed requirements. Scenario C represents a resource-constrained SME setting, where smaller batch size, cost containment and waste avoidance become more relevant.

In addition, one-way sensitivity analyses were performed on the economic parameters most likely to affect model selection. The hourly cost of downtime was varied in Scenario A to test the stability of the throughput-driven ranking. The quality-to-scrap cost ratio C_Q/C_{scrap} was varied across all three scenarios to examine the relative importance of missed Cracked defects compared with unnecessary rejection. Finally, the safety-related missed-defect cost C_S was varied in Scenarios B and C to assess whether the preference for risk-oriented profiles depended on a narrow calibration of the safety-risk weight. These analyses are reported jointly in Fig. 5.

4. Experimental results

This section presents the empirical findings in two steps. First, it compares the technical performance of the four detection models (YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN) across the Nominal and Synthetic Perturbation dataset conditions. Second, it interprets these results through the scenario-based managerial framework, focusing on cost exposure, throughput feasibility, and waste-related implications under the assumptions defined in Section 3.4.

4.1. Performance and computational efficiency

Table 7 summarises the technical performance of YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN under Nominal and Synthetic Perturbation conditions. The three YOLO architectures achieved consistently high detection performance, with mAP@50 values ranging from 0.963 to 0.976 and mAP@50-95 values ranging from 0.941 to 0.951 across conditions. By contrast, Faster R-CNN exhibited lower aggregate detection performance, particularly in terms of mAP@50-95.

The main operational variations among models are therefore driven less by marginal differences in aggregate detection scores and more by latency and class-dependent error profiles. YOLOv11 provided the most favourable latency profile, with total latency decreasing from 18.4 ± 0.6 ms under Nominal conditions to 14.6 ± 0.1 ms under Perturbation, corresponding to approximately 54.3 and 68.5 FPS, respectively. YOLOv5 and YOLOv8 showed comparable latency under Perturbation, with 27.4 ± 1.0 ms and 24.8 ± 0.8 ms, respectively, whereas Faster R-CNN remained the slowest model. Latency results indicate that YOLOv11 achieves the lowest end-to-end inference time, consistent with a more efficient post-processing stage.

4.2. Diagnostic analysis of classification challenges

The normalised confusion matrices (Fig. 3) reveal distinct FP/FN profiles with direct implications for cost-, waste- and risk-sensitive model selection.

Across the three YOLO architectures, Cracked defects were detected more consistently than Over-baked defects. The Cracked class shows high diagonal values, ranging around 0.91, whereas the Over-baked class remains more frequently confused with Normal products. In particular, the proportion of Over-baked samples predicted as Normal is 0.29 for YOLOv5 and 0.26 for both YOLOv8 and YOLOv11. This pattern suggests that colour- and texture-related deviations associated with over-baking are more difficult to separate from acceptable product variability than structural defects such as cracking. Table 8 illustrates qualitative examples of easy, difficult, and failure cases under Synthetic Perturbation conditions.

Although Faster R-CNN correctly identifies most Normal products and detects Cracked defects with reasonable consistency, it exhibits a weaker diagnostic profile for Over-baked samples, with 0.35 of Over-baked instances predicted as Normal and a lower diagonal value of 0.65 for the Over-baked class. Its Cracked diagonal value is also lower than that of the YOLO models, at 0.82.

4.3. Scenario analysis

To translate the technical results into decision-support evidence, the four models were evaluated across the three manufacturing scenarios defined in Table 6 by integrating Risk-Weighted TCQ, TE and WRI. The economic simulation uses the class-level error profiles obtained under Synthetic Perturbation conditions and the mean total latency values reported in Table 7.

4.3.1. Scenario A: commodity production

In Scenario A, as visualised in Fig. 4 (Panel A), YOLOv11 achieves the lowest TCQ, equal to 2698 €/batch, followed by YOLOv8 with 2810 €/batch, YOLOv5 with 3214 €/batch, and Faster R-CNN 3894 €/batch. Although YOLOv8 shows a favourable quality-risk profile, its effective speed of approximately 40.3 FPS generates a downtime penalty of about 133 €/batch. YOLOv11, being the only model with 100% TE, avoids this penalty entirely. These results indicate that, in high-throughput commodity production, marginal differences in risk-weighted defect costs can be outweighed by the economic value of maintaining line speed.

4.3.2. Scenario B: Brand Protection

In Scenario B (Fig. 4, Panel B), downtime does not contribute to TCQ because all models exceed the target speed. YOLOv8 yields the lowest TCQ, equal to 3446 €/batch, followed closely by YOLOv11. YOLOv5 reaches 3898 €/batch, while Faster R-CNN records the highest TCQ. The advantage of YOLOv8 is mainly attributable to its lower quality-risk contribution for missed Cracked defects, while maintaining the same safety-risk contribution as YOLOv11. The small difference between YOLOv8 and YOLOv11 suggests that both architectures are viable, but YOLOv8 provides the most favourable risk-weighted profile under the specific case study.

4.3.3. Scenario C: SME

In Scenario C (Fig. 4, Panel C), YOLOv8 achieves the lowest TCQ, equal to 866 €/batch, but its advantage over YOLOv11 is marginal, with YOLOv11 reaching 871 €/batch. YOLOv5 records 977 €/batch, whereas Faster R-CNN reaches 1200 €/batch. This limited difference suggests that the ranking between YOLOv8 and YOLOv11 should be interpreted cautiously, whereas YOLOv5 is not economically preferred in TCQ terms. Its relevance lies instead in its lower FP profile and therefore in its

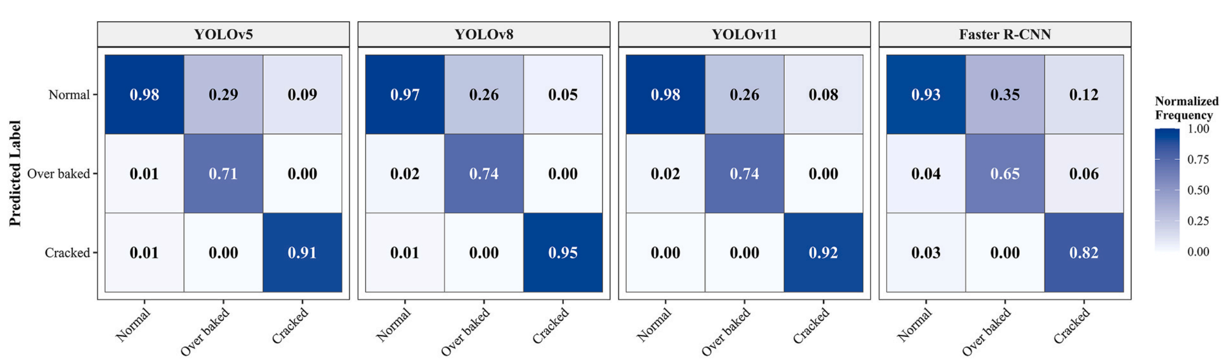
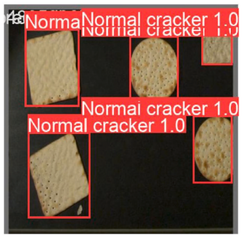
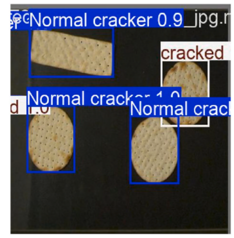
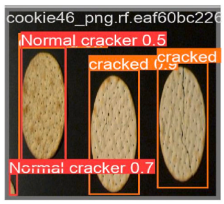
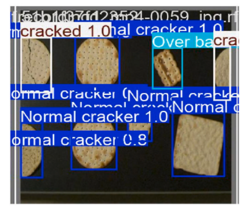
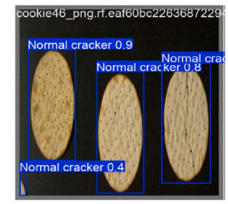
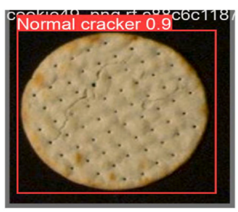
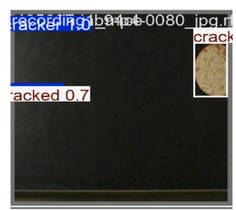
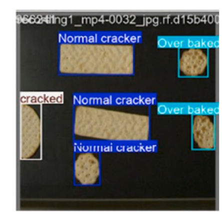


Fig. 3. Normalised confusion matrices under Synthetic Perturbation conditions for YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN. Values are averaged over three independent seeds.

Table 8
Qualitative examples of YOLOv5, YOLOv8 and YOLOv11 predictions under Synthetic Perturbation conditions. The figures show representative "easy", "difficult" and "failure" cases for each model. Differences in FP/FN profiles illustrate how the three architectures trade off.

Case	YOLOv5	YOLOv8	YOLOv11
Easy	 TP: 4 FP:0 FN:0	 TP: 3 FP:1 FN:0	 TP:1 FP:0 FN:0
Difficult	 TP: 3 FP:0 FN:0	 TP: 7 FP: 0 FN:1	 TP:2 FP:1 FN:0
Failure	 TP:0 FP:1 FN:0	 TP: 0 FP: 0 FN:1	 TP: 5 FP:1 FN:1

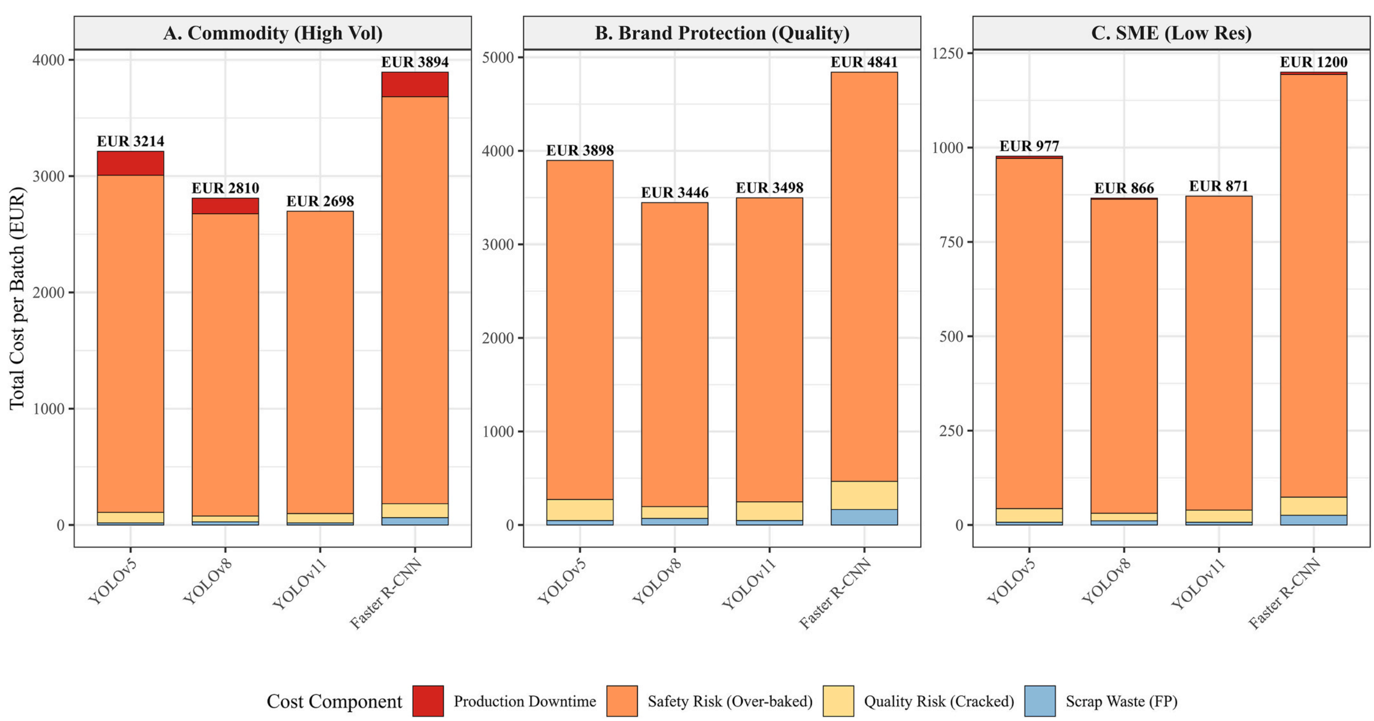


Fig. 4. Panel A shows the high-volume commodity production scenario, Panel B the brand-protection scenario, and Panel C the resource-constrained SME scenario. Bars report the TCQ decomposition by cost component for each model under the adopted scenario assumptions.

stronger waste-avoidance behaviour.

4.3.4. Sensitivity analyses of key cost assumptions

Because the managerial ranking depends on scenario-dependent decision weights, additional one-way sensitivity analyses were

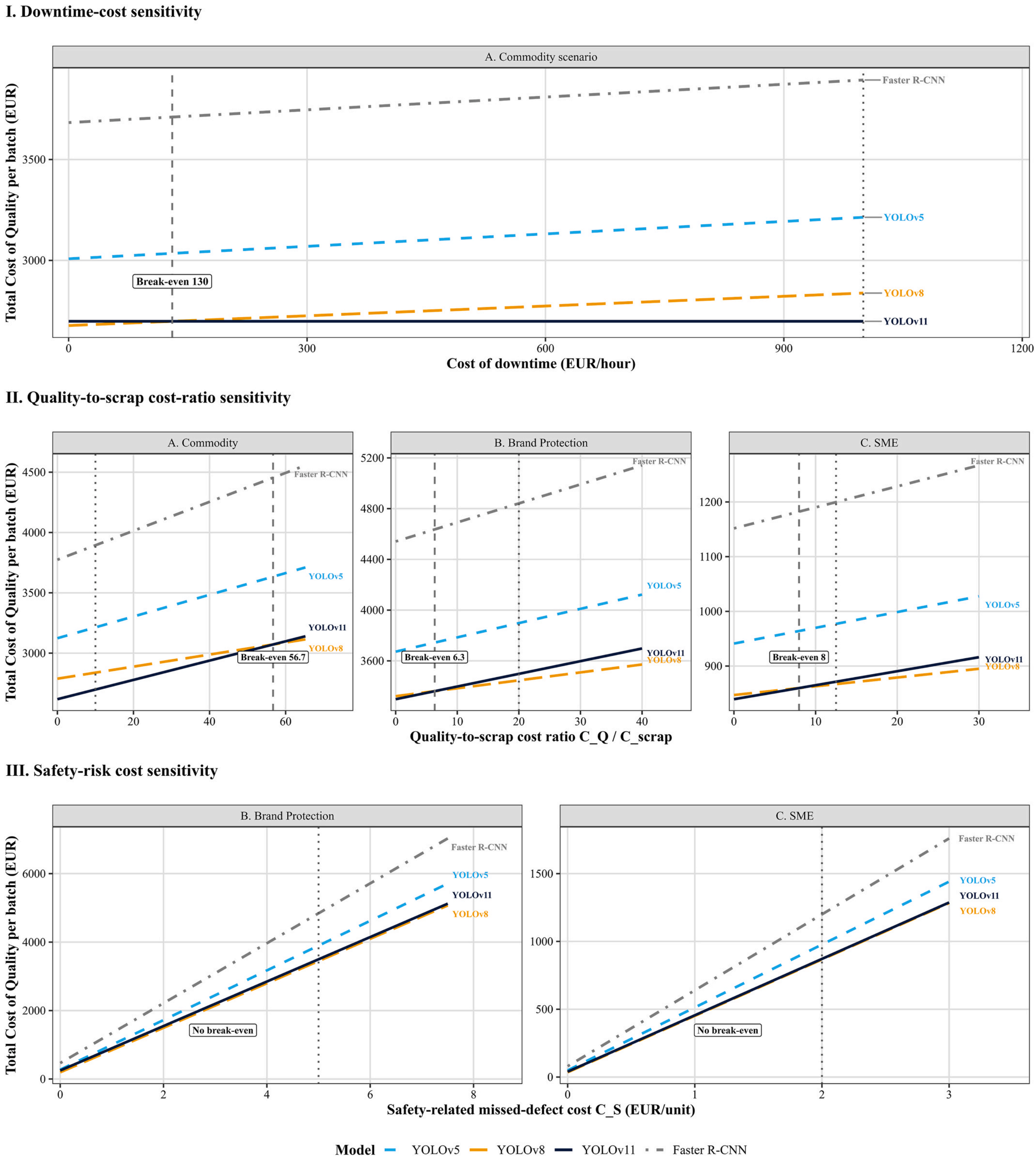


Fig. 5. One-way sensitivity analyses on the main economic assumptions of the Risk-Weighted TCQ model. Panel I reports the effect of the hourly cost of downtime in the Commodity scenario. Panel II reports the effect of the quality-to-scrap cost ratio C_Q/C_{scrap}. Panel III reports the effect of the safety-related missed-defect cost C_S. Dashed vertical lines indicate break-even points between YOLOv8 and YOLOv11 where present; dotted vertical lines indicate the baseline values adopted in Table 6.

performed on the cost parameters most likely to affect model selection. Three parameters were examined: the hourly cost of downtime, the quality-to-scrap cost ratio, and the unit cost assigned to missed risk-bearing defects.

First, the robustness of the Commodity scenario was tested by varying the $C_{downtime}$ parameter. As shown in Fig. 5 Panel I, YOLOv8 is slightly preferable only when the cost of downtime is very low. The break-even point between YOLOv8 and YOLOv11 occurs at approximately 130 €/hour.

Second, as reported in Fig. 5 Panel II, a complementary sensitivity analysis was performed on the quality-to-scrap cost ratio, defined as C_Q/C_{scrap}. This ratio captures the relative economic importance assigned to missed aesthetic or quality defects compared with unnecessary rejection of compliant products. The resulting break-even points differ substantially across the three scenarios. In Scenario A, the break-even ratio between YOLOv11 and YOLOv8 is 56.7. This confirms that YOLOv11 remains preferable in throughput-dominated production unless the relative penalty for missed Cracked defects becomes extremely high. In Scenario B, the break-even ratio is 6.3; therefore, YOLOv8 is robustly preferred in the Brand Protection scenario. In Scenario C, the break-even ratio is 8.0, again supporting the slight economic preference for YOLOv8 in the SME scenario, although the margin over YOLOv11 remains limited.

Finally, the sensitivity analysis on the unit cost assigned to missed risk-bearing defects, C_S, shows no break-even between YOLOv8 and YOLOv11 in the explored ranges for Scenarios B and C (Fig. 5 Panel III). In both cases, YOLOv8 remains the lowest-cost option, while YOLOv11 follows closely. Faster R-CNN remains consistently penalised by its higher missed-risk profile, while YOLOv5 remains less competitive because of its higher safety-risk contribution.

4.3.5. *Trade-off analysis*

The scenario-based results were further synthesised through a radar chart focusing on the most throughput-constrained setting, namely Scenario A (Fig. 6). The chart compares the four models along four managerial dimensions: TE, TCQ, risk containment, and WRI.

YOLOv11 shows the most balanced profile. It reaches the highest score on the TE axis and performs strongly in terms of TCQ, confirming that its economic advantage is mainly driven by its ability to sustain the required line speed without generating downtime losses.

YOLOv8 presents a different profile. It is less favourable than YOLOv11 on TE, but it remains highly competitive on risk containment and TCQ performance. This confirms its role as a quality-oriented alternative when line-speed constraints are less stringent, as observed in Scenarios B and C. YOLOv5 occupies an intermediate position: it does not dominate any dimension, but it remains a viable legacy or low-complexity option when deployment simplicity and FP control are prioritised. Faster R-CNN shows the weakest overall profile, confirming that the non-YOLO baseline does not provide a competitive trade-off under the adopted assumptions.

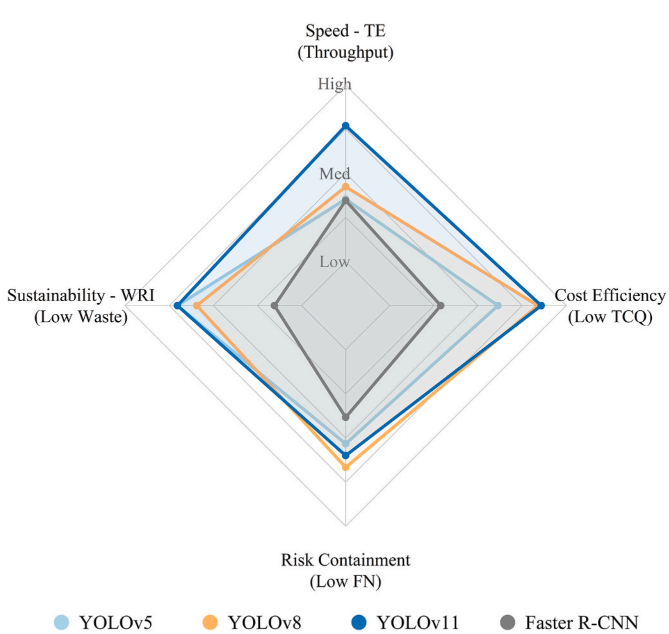


Fig. 6. Strategic trade-off analysis for Scenario A. The radar chart compares YOLOv5, YOLOv8, YOLOv11 and Faster R-CNN across four normalised managerial dimensions. Higher values indicate more favourable performance.

5. Discussion

The results confirm that the value of an AI-based inspection model cannot be inferred from detection metrics alone. Across the four architectures, the economic ranking changed substantially once latency, class-specific errors and production constraints were introduced. This indicates that the transferable element of the study is primarily the decision logic, rather than the numerical ranking observed in the bakery case. The latter depends on defect prevalence, cost weights, line-speed requirements and deployment conditions, consistent with recent calls to incorporate economic and organisational considerations into the evaluation of digital inspection systems (Khan et al., 2020; Olakanmi et al., 2023; Perez Colo et al., 2023). This finding aligns with the Quality 4.0 view that digital tools create value through context-sensitive, data-driven decision-making (Sony and Naik, 2020; Talaie et al., 2024), and operationalises the bottom-up framework proposed by Piovano et al. (2026) by translating shop-floor constraints into an explicit model-selection rationale.

5.1. *Latency as a cost-of-quality component*

The Commodity scenario highlights that latency is not a secondary computational attribute, but a cost-of-quality component when inspection speed affects line continuity. Under the 50 FPS requirement, YOLOv11 achieved the lowest TCQ because it was the only model able to exceed the target throughput with sufficient margin. This extends the economic logic discussed by Verna et al. (2022), showing that in Industry 4.0 settings the return on inspection systems must account for the temporal cost of the algorithm itself. If inference time forces buffering, re-routing or line slowdown, the inspection system becomes a source of quality cost rather than a neutral monitoring layer.

5.2. *Risk weighting and the role of class-specific errors*

The interpretation changes when throughput is no longer the dominant constraint. In the Brand Protection scenario, all models exceed the target line speed, and the ranking is driven mainly by the weighted consequences of missed defects. Under these assumptions, YOLOv8 provides the lowest TCQ, although YOLOv11 remains close. As highlighted by Piovano et al. (2026), food safety is a non-negotiable stakeholder requirement that must drive the adoption of Industry 4.0 technologies. This result illustrates why class-level diagnostics are necessary in food inspection. In premium segments, visual consistency is a direct quality indicator and a primary driver of consumer rejection (Kumar and Thankachan, 2025); model selection is therefore determined by the tolerance for specific defect categories rather than by aggregate accuracy.

5.3. *SME settings and the limits of a "best model" interpretation*

Scenario C (SME) provides a more nuanced result. YOLOv8 achieves the lowest TCQ, but its advantage over YOLOv11 is small. In this case, the lower batch size and lower downtime cost reduce the economic importance of latency, while the weighted error profile remains relevant. YOLOv11 therefore loses part of the advantage observed in the high-volume scenario, but still offers a stronger throughput margin. YOLOv5 relevance is associated with cases where FP minimisation, deployment familiarity or compatibility with legacy hardware are prioritised over the best overall economic score. This is particularly important for SMEs, where the selection of an AI model may be constrained by available computing infrastructure, staff expertise and maintenance requirements.

5.4. *Implications for quality 4.0 model selection*

The three scenarios show that model-selection criteria shift with production priorities: latency costs dominate in high-throughput commodity production, class-level error costs dominate in risk-sensitive premium settings, and resource constraints become binding in SME contexts. The value of automated inspection thus depends not only on defect detection capability, but on the actionability of inspection outputs within the production decision loop (Khan et al., 2025).

Table 9 summarises this interpretation by linking production contexts to model-selection priorities. The matrix should therefore not be read as a universal ranking of architectures, but as a decision aid for aligning inspection technology with manufacturing strategy. In other bakery lines or food categories, the same logic would require recalibration of defect severities, takt-time constraints, hardware assumptions and cost-of-quality parameters.

6. Conclusion

This study proposes and applies a decision-oriented framework for selecting AI-based visual inspection models in food manufacturing. The framework combines conventional detection metrics with Risk-Weighted TCQ, TE and WRI in order to evaluate whether a model is suitable for a given production context. The empirical bakery case shows that higher aggregate detection performance does not automatically translate into higher industrial value. Once latency and asymmetric error costs are considered, the preferred model changes according to the operating scenario.

The comparison reveals that model suitability is scenario-dependent rather than determined by aggregate accuracy. YOLOv11 offers the strongest profile in throughput-constrained commodity production; YOLOv8 is preferable when class-level missed-defect costs outweigh latency penalties; and YOLOv5 remains viable when false-positive control or legacy compatibility are prioritised. Faster R-CNN, included as a non-YOLO baseline, was consistently dominated in the investigated case, showing that the same cost-quality-throughput logic can also be applied to architectures outside the YOLO family without changing the managerial interpretation of the trade-offs.

6.1. Theoretical contributions

The study contributes to the Quality 4.0 and cost-of-quality literature in three respects. First, it extends the CoQ-PAF framework (Verna et al., 2022; Plunkett and Dale, 1987) by quantifying the cost of latency as a distinct internal failure component. Second, it operationalises severity-differentiated FN classification within the ZDM cost-modelling tradition (Psarommatis et al., 2020), moving beyond binary TP/FP evaluation. Third, it provides an empirically grounded comparison protocol for food MV model selection under explicit production constraints, contributing a replicable evidence base for Quality 4.0 adoption decisions in food manufacturing.

6.2. Managerial implications

For practitioners, the main implication is that AI adoption in food inspection should start from the production constraint rather than from an aggregate benchmark ranking. A plant operating close to its line-speed limit should prioritise latency and throughput margin. A premium or risk-sensitive producer should place greater emphasis on the class-specific distribution of false negatives. A resource-constrained SME should evaluate not only TCQ, but also deployment complexity, hardware availability and the cost of unnecessary scrap.

The proposed framework can support this decision process by making the trade-offs explicit. It provides a shared evaluation protocol for quality engineers, food technologists and data scientists to compare inspection architectures in terms of cost exposure, throughput feasibility and waste implications, dimensions that are particularly binding in SME contexts where infrastructure, staff expertise and deployment complexity constrain technology adoption.

6.3. Limitations and future work

The study has some limitations that should be considered when interpreting the results. First, the economic parameters were defined as scenario-dependent decision weights rather than as audited plant-level accounting values. The TCQ values should therefore be interpreted as comparative indicators for model selection, not as direct estimates of the actual cost structure of a specific bakery plant. Second, the stress condition was generated through controlled synthetic perturbations. This provides a reproducible way to compare model behaviour under degraded visual conditions, but it does not fully reproduce the complexity of real production environments, where illumination drift, product overlap, belt contamination, camera vibration and seasonal process variability may jointly affect acquisition quality, model stability and rejection decisions. Third, the empirical evidence is limited to one bakery-defect dataset and to a three-class inspection problem. The Overbaked class was used as a visual proxy for severe thermal-processing non-conformities and should not be interpreted as a direct measurement of food safety risk.

Future work should validate the framework on plant-level data and real in-line acquisition systems, including longer production runs, externally collected datasets and shadow-mode testing in which model predictions are compared with existing inspection decisions before closed-loop rejection is activated. Further research should also extend the comparison to other detector families and to additional food products with different defect structures and cost asymmetries. Finally, the proposed framework could be integrated with digital twins, MES-connected adaptive thresholds and human-in-the-loop inspection strategies, so that model selection and inspection policies can be updated as process conditions and business priorities change.

Table 9
Cost-Quality-Throughput decision matrix for aligning AI model selection with manufacturing priorities.

Production Context	Model-selection priority	Most suitable profile in the present case	Dominant constraint	Quality 4.0 interpretation
Efficiency-driven mass production, e.g. standard biscuits or high-volume crackers	Maintain line continuity and avoid latency-induced downtime	YOLOv11	Takt-time compatibility and throughput margin	Inspection must not become the bottleneck; latency is part of the cost-of-quality structure.
Risk-sensitive premium production, e.g. export bakery or brand-protected products	Reduce weighted consequences of missed defects	YOLOv8	External failure exposure and class-specific false negatives	Model selection is driven by risk containment rather than by aggregate accuracy alone.
Resource-constrained SME production, e.g. small batches or limited hardware availability	Balance TCQ, deployment simplicity and waste avoidance	YOLOv8 or YOLOv11 depending on constraint; YOLOv5 as low-complexity alternative	Limited hardware, staff expertise and tolerance for scrap	The preferred model is context-dependent; “good-enough” deployment may be more relevant than peak benchmark performance.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT solely to support language editing and improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This paper is part of the project NODES which has received funding from the MUR-M4C2 1.5 of PNRR funded by the European Union – NextGenerationEU (Grant agreement no. ECS00000036).

CRediT authorship contribution statement

Alberto Piovano: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Elisa Verna:** Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing. **Nazarena Cela:** Conceptualization, Methodology, Writing – review & editing. **Luisa Torri:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Gianfranco Genta:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Maurizio Galetto:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Ms. Lucia Ruffati for her valuable contributions to this work.

Appendix

Table A1
Main hyperparameters used for training the YOLO models

Hyper-parameter	Value	Description	Reference
Learning Rate	0.01	Default YOLO learning rate	Li et al. (2024)
Optimizer	Adam	Adaptive weight optimization	(Ultralytics)
Batch size	8	Tuned to GPU capacity	(Ultralytics)
Epochs	100	Ensures convergence without underfitting	Drogalis et al. (2024)
Weight Decay	0.0005	Regularization term	Li et al. (2024)
IoU threshold	0.5	Bounding-box overlap criterion	Li et al. (2024)

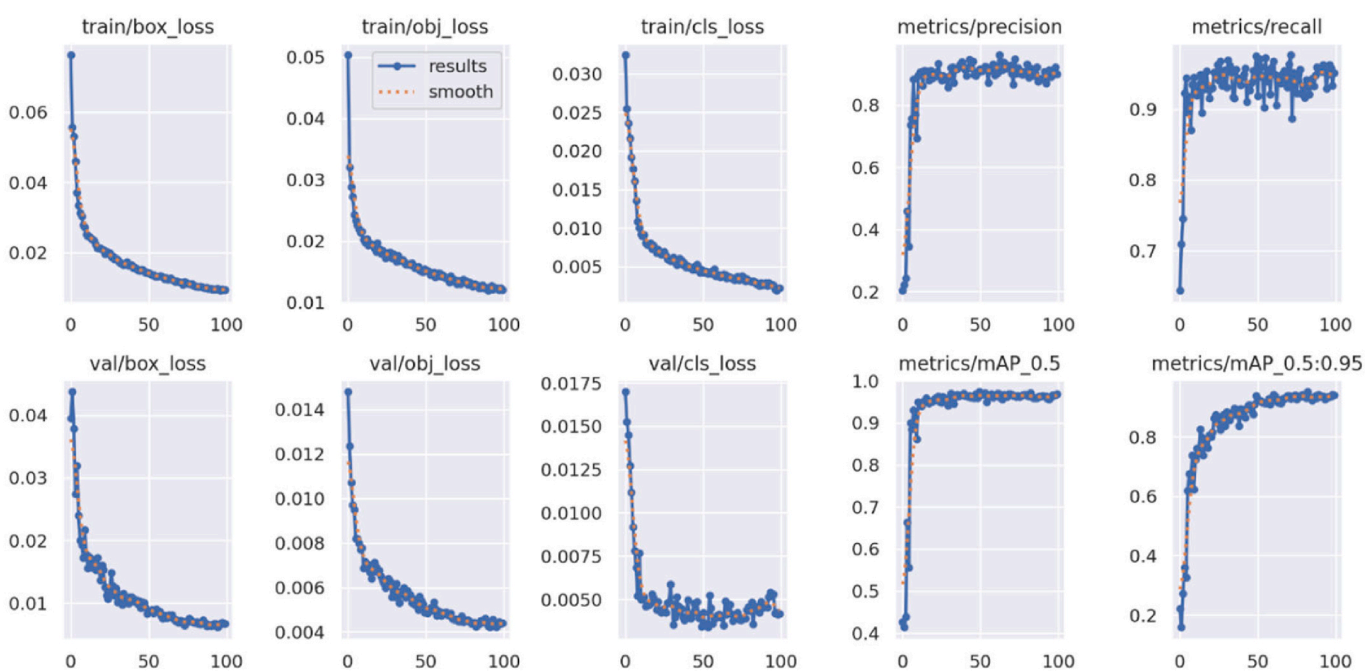


Fig. A1. Training and validation loss curves under synthetic perturbation training scenario.

Table A2
Class-level performance metrics for Normal, Cracked and Over-baked classes. Precision, recall, mAP@50 and mAP@50–95 are reported for each defect class, model and dataset condition, using the first reference run (seed 101) for diagnostic comparison. The table highlights the relative difficulty of correctly detecting Over-baked products compared with Normal and Cracked classes.

Model	Scenario	Class	Precision	Recall	mAP@50	mAP50-95
YOLOv5	Nominal	Normal	0.923	0.991	0.986	0.956
		Cracked	0.884	0.978	0.992	0.990
		Over baked	0.903	0.934	0.948	0.928
	Perturbation	Normal	0.934	0.973	0.987	0.965
		Cracked	0.956	0.938	0.987	0.981
		Over baked	0.867	0.9	0.942	0.912
YOLOv8	Nominal	Normal	0.908	0.969	0.985	0.969
		Cracked	0.894	0.978	0.990	0.982
		Over baked	0.841	0.880	0.936	0.909
	Perturbation	Normal	0.960	0.943	0.988	0.970
		Cracked	0.939	1.000	0.992	0.976
		Over baked	0.885	0.862	0.966	0.940
YOLOv11	Nominal	Normal	0.927	0.973	0.988	0.969
		Cracked	0.978	0.957	0.987	0.985
		Over baked	0.825	0.933	0.938	0.922
	Perturbation	Normal	0.954	0.947	0.989	0.967
		Cracked	0.986	0.957	0.99	0.989
		Over baked	0.896	0.901	0.955	0.926
Faster R-CNN	Nominal	Normal	0.889	0.945	0.938	0.786
		Cracked	0.710	0.846	0.863	0.788
		Over baked	0.750	0.833	0.801	0.704
	Perturbation	Normal	0.815	0.937	0.946	0.770
		Cracked	0.600	0.923	0.928	0.724
		Over baked	0.619	0.722	0.773	0.643

Table A3
Object-level class distribution across dataset conditions.

Dataset condition	Split	Images	Normal instances	Cracked instances	Over-baked instances
Nominal	Train	537	1657	361	233
Perturbation	Train	924	2805	618	396

Data availability

Data will be made available on request.

References

Abdanan Mehdizadeh, S., 2022. Machine vision based intelligent oven for baking inspection of cupcake: design and implementation. Mechatronics 82, 102746. https://doi.org/10.1016/j.mechatronics.2022.102746.

Agrawal, K., Goktas, P., Holtkemper, M., Beecks, C., Kumar, N., 2025. AI-driven transformation in food manufacturing: a pathway to sustainable efficiency and quality assurance. Front. Nutr. 12, 1553942. https://doi.org/10.3389/fnut.2025.1553942.

Babor, M., Paquet-Durand, O., Kohlus, R., Hitzmann, B., 2023. Modeling and optimization of bakery production scheduling to minimize makespan and oven idle time. Sci. Rep. 13, 235. https://doi.org/10.1038/s41598-022-26866-9.

Broyart, B., Trystram, G., 2002. Modelling heat and mass transfer during the continuous baking of biscuits. J. Food Eng. 51 (1), 47–57. https://doi.org/10.1016/S0260-8774(01)00036-X.

Cai, Z., Zhou, K., Liao, Z., 2025. A systematic review of YOLO-based object detection in medical imaging: advances, challenges, and future directions. Comput. Mater. Contin. 85 (2), 2255–2303. https://doi.org/10.32604/CMC.2025.067994/S1.

Capuano, E., Ferrigno, A., Acampa, I., Ait-Ameur, L., Fogliano, V., 2008. Characterization of the Maillard reaction in bread crisps. Eur. Food Res. Technol. 228 (2), 311–319. https://doi.org/10.1007/s00217-008-0936-5.

Chen, J., Wu, Y., Jiang, X., 2022. Research on consumer purchasing channel choice based on product tolerance: the mediating role of rationalization. Front. Psychol. 13, 823470. https://doi.org/10.3389/fpsyg.2022.823470.

Chiarini, A., Kumar, M., 2022. What is quality 4.0? An exploratory sequential mixed methods study of Italian manufacturing companies. Int. J. Prod. Res. 60 (16), 4890–4910. https://doi.org/10.1080/00207543.2021.1942285.

Dikici, S., Robinson, R.J., 2024. Automated defect detection using image recognition in manufacturing. J. Data Sci. Intell. Syst. https://doi.org/10.47852/BONVIEWJDSIS42023833.

Drogalis, C., Zampino, C., Chauhan, V., 2024. Food quality inspection and sorting using machine vision, machine learning and robotics. In: Proceedings of the ASME International Mechanical Engineering Congress and Exposition, vol. 3. IMECE. https://doi.org/10.1115/IMECE2023-113496.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. 88 (2), 303–338. https://doi.org/10.1007/S11263-009-0275-4.

Farooq, M.A., Kirchain, R., Novoa, H., Araujo, A., 2017. Cost of quality: evaluating cost-quality trade-offs for inspection strategies of manufacturing processes. Int. J. Prod. Econ. 188, 156–166. https://doi.org/10.1016/j.ijpe.2017.03.019.

Feng, Zhao L., Xia, Y., Dai, L., 2025. Unveiling the dual impacts of the aesthetic deficiency of foods on consumers' purchase intentions. Sci. Rep. 15 (1), 11218. https://doi.org/10.1038/s41598-025-92352-7.

Gökmen, V., Mogol, B.A., 2010. Computer vision-based image analysis for rapid detection of acrylamide in heated foods. Qual. Assur. Saf. Crops Foods 2 (4), 203–207. https://doi.org/10.1111/j.1757-837X.2010.00072.x.

Gomez, C.B., Marks, B.P., 2020. Monetizing the impact of food safety recalls on the low-moisture food industry. J. Food Prot. 83 (5), 829–835. https://doi.org/10.4315/JFP-19-553.

Görgülü, A., 2025. Real-time quality analysis of baked goods using advanced technologies. J. Food Eng. 388, 112359. https://doi.org/10.1016/j.jfoodeng.2024.112359.

Hanhirova, J., Kämäräinen, T., Seppälä, S., Siekkinen, M., Hirvisalo, V., Ylä-Jääski, A., 2018. Latency and throughput characterization of convolutional neural networks for mobile computer vision. In: Proceedings of the 9th ACM Multimedia Systems Conference, pp. 204–215. https://doi.org/10.1145/3204949.3204975. MMSys 2018.

Hutmacher, F., 2019. Why is there so much more research on vision than on any other sensory modality? Front. Psychol. 10, 481030. https://doi.org/10.3389/fpsyg.2019.02246.

Khan, A., Munir, M.T., Yu, W., Young, B.R., 2020. A review towards hyperspectral imaging for real-time quality control of food products with an illustrative case study of milk powder production. Food Bioprocess Technol. 13 (5), 739–752. https://doi.org/10.1007/S11947-020-02433-W.

Khan, S., Mehmood, S., Khan, S.U., 2025. Navigating innovation in the age of AI: how generative AI and innovation influence organisational performance in the

manufacturing sector. J. Manuf. Technol. Manag. 36 (3), 597–620. https://doi.org/10.1108/JMTM-06-2024-0302.
Kılcı, O., Koklu, M., 2025. Automated classification of biscuit quality using YOLOv8 models in food industry. Food Anal. Methods 18 (5), 815–829. https://doi.org/10.1007/s12161-025-02755-5.
Kumar, S., Thankachan, J., 2025. Aesthetic of consumption: a review of sensory branding and visual semiotics in consumer experience. Adv. Consum. Res. 2, 1068–1083.
Lee, J., Kim, Y., Kim, S., 2023. The study of an adaptive bread maker using machine learning. Foods 12 (22), 4160. https://doi.org/10.3390/foods12224160.
Lehn, F., Schmidt, T., 2022. Sustainability assessment of food-waste-reduction measures by converting surplus food into processed food products for human consumption. Sustainability 15 (1), 635. https://doi.org/10.3390/SU15010635.
Li, X., Xue, S., Li, Z., Fang, X., Zhu, T., Ni, C., 2024. A candy defect detection method based on StyleGAN2 and improved YOLOv7 for imbalanced data. Foods 13 (20), 3343. https://doi.org/10.3390/FOODS13203343.
Mannaro, K., Baire, M., Fanti, A., Lodi, M.B., Didaci, L., Fedeli, A., Cocco, L., Randazzo, A., Mazzarella, G., Fumera, G., 2022. A robust SVM color-based food segmentation algorithm for the production process of a traditional Carasau bread. IEEE Access 10, 15359–15377. https://doi.org/10.1109/ACCESS.2022.3147206.
Mogol, B.A., Gökmen, V., 2014. Computer vision-based analysis of foods: a non-destructive colour measurement tool to monitor quality and safety. J. Sci. Food Agric. 94 (7), 1259–1263. https://doi.org/10.1002/jsfa.6500.
Muffato Reis, A., Verna, E., Costa, L., Sousa, S.D., Galetto, M., 2025. A scenario-based framework for strategic inspection decision-making in high-volume production environments. Int. J. Qual. Reliab. Manag. 42 (6), 1822–1850. https://doi.org/10.1108/IJQRM-03-2024-0100.
Mumuni, A., Mumuni, F., 2022. Data augmentation: a comprehensive survey of modern approaches. Array 16, 100258. https://doi.org/10.1016/j.array.2022.100258.
Munoz, A.M., 2002. Sensory evaluation in quality control: an overview, new developments and future opportunities. Food Qual. Prefer. 13 (6), 329–339. https://doi.org/10.1016/S0950-3293(02)00014-9.
Nakajima, S., 1988. Introduction to TPM: Total Productive Maintenance. Productivity Press, Cambridge, MA.
Ogidi, O.I., Wenapere, C.M., Chukwunonso, O.A., 2025. Enhancing food safety and quality control with computer vision systems. In: Bora, D., Bania, R. (Eds.), Computer Vision Techniques for Agricultural Advancements. IGI Global Scientific Publishing, pp. 51–88. https://doi.org/10.4018/979-8-3693-8019-2.ch003.
Olakanmi, S.J., Jayas, D.S., Paliwal, J., 2023. Applications of imaging systems for the assessment of quality characteristics of bread and other baked goods: a review. Compr. Rev. Food Sci. Food Saf. 22 (3), 1817–1838. https://doi.org/10.1111/1541-4337.13131.
Perez Colo, I., Saavedra Sueldo, C., De Paula, M., Acosta, G.G., 2023. Intelligent approach for the industrialization of deep learning solutions applied to fault detection. Expert Syst. Appl. 233, 120959. https://doi.org/10.1016/j.eswa.2023.120959.
Piovano, A., Verna, E., Genta, G., Galetto, M., 2026. Industry 4.0 enablers in food industry: a framework based on stakeholder challenges in quality and sustainability. Int. J. Qual. Reliab. Manag. 43 (2), 349–372. https://doi.org/10.1108/IJQRM-04-2025-0120/1300549.
Plunkett, J.J., Dale, B.G., 1987. A review of the literature on quality-related costs. Int. J. Qual. Reliab. Manag. 4 (1), 40–52. https://doi.org/10.1108/EB002874.
Porter, L.J., Rayner, P., 1992. Quality costing for total quality management. Int. J. Prod. Econ. 27 (1), 69–81. https://doi.org/10.1016/0925-5273(92)90127-S.
Psarommatis, F., May, G., Dreyfus, P.A., Kiritsis, D., 2020. Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. Int. J. Prod. Res. 58 (1), 1–17. https://doi.org/10.1080/00207543.2019.1605228.
Psomas, E., Dimitrantzou, C., Vouzas, F., Bouranta, N., 2018. Cost of quality measurement in food manufacturing companies: the Greek case. Int. J. Product. Perform. Manag. 67 (9), 1882–1900. https://doi.org/10.1108/IJPPM-12-2017-0325.
Purlis, E., 2010. Browning development in bakery products – a review. J. Food Eng. 99 (3), 239–249. https://doi.org/10.1016/j.jfoodeng.2010.03.008.
Raak, N., Symmank, C., Zahn, S., Aschemann-Witzel, J., Rohm, H., 2017. Processing- and product-related causes for food waste and implications for the food supply chain. Waste Manag. 61, 461–472. https://doi.org/10.1016/j.wasman.2016.12.027.
Radziwill, N., 2020. Why digital transformations fail: the surprising disciplines of how to take off and stay ahead. Qual. Manag. J. 27 (4), 242. https://doi.org/10.1080/10686967.2020.1812988.
Reichenstein, T., Raffin, T., Sand, C., Franke, J., 2022. Implementation of machine vision based quality inspection in production: an approach for the accelerated execution of case studies. Proced. CIRP 112, 596–601. https://doi.org/10.1016/j.procir.2022.09.058.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
Rezagholi, F., Hesarinejad, M.A., 2017. Integration of fuzzy logic and computer vision in intelligent quality control of celiac-friendly products. Procedia Comput. Sci. 120, 325–332. https://doi.org/10.1016/j.procs.2017.11.246.
Rohan Joel, J., Benjamin, R.M., 2024. Bread quality assessment using deep learning with shape and volume metrics. In: 10th International Conference on Advanced Computing and Communication Systems, pp. 2125–2130. https://doi.org/10.1109/ICACCS60874.2024.10717194. ICACCS 2024.
Senni, L., Ricci, M., Palazzi, A., Burrascano, P., Pennisi, P., Ghirelli, F., 2014. On-line automatic detection of foreign bodies in biscuits by infrared thermography and image processing. J. Food Eng. 128, 146–156. https://doi.org/10.1016/j.jfoodeng.2013.12.016.
Sony, M., Naik, S., 2020. Industry 4.0 integration with socio-technical systems theory: a systematic review and proposed theoretical model. Technol. Soc. 61, 101248. https://doi.org/10.1016/j.techsoc.2020.101248.
Srivastava, S., Boyat, S., Sadistap, S., 2014. A robust machine vision algorithm development for quality parameters extraction of circular biscuits and cookies digital images. J. Food Process. 2014, 376360. https://doi.org/10.1155/2014/376360.
Talaie, H., Ziaeian, M., Malekinejad, P., 2024. Towards Quality 4.0 in home appliances: definitions, deployment scenarios, and future perspectives. J. Manuf. Technol. Manag. 35 (7), 1416–1438. https://doi.org/10.1108/JMTM-02-2023-0044.
Tantiphanwadi, P., Malithong, K., 2022. Bread browning stage classification model using VGG-16 transfer learning and fine-tuning with small training dataset. Eng. J. 26 (11), 1–14. https://doi.org/10.4186/EJ.2022.26.11.1.
Terven, J., Córdova-Esparza, D.M., Romero-González, J.A., 2023. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr. 5 (4), 1680–1716. https://doi.org/10.3390/make5040083.
Tutuncu, K., Yasin, E.T., Koklu, M., 2023. Enhancing quality control: defect state classification of Taralli biscuits with MobileNet-v2 and DenseNet-201. In: Proceedings of the IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. IDAACS, pp. 718–723. https://doi.org/10.1109/IDAACS58523.2023.10348851.
Verna, E., Genta, G., Galetto, M., Franceschini, F., 2022. Economic impact of quality inspection in manufacturing: a proposal for a novel cost modeling. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. 236 (11), 1508–1517. https://doi.org/10.1177/09544054221078090.
Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., Wellbrock, W., 2019. Deep learning for industrial computer vision quality control in the printing industry 4.0. Sensors 19 (18), 3987. https://doi.org/10.3390/S19183987.
Wuest, T., Weimer, D., Irgens, C., Thoben, K.D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. Prod. Manuf. Res. 4 (1), 23–45. https://doi.org/10.1080/21693277.2016.1192517.
Xiao, Z., Wang, J., Han, L., Guo, S., Cui, Q., 2022. Application of machine vision system in food detection. Front. Nutr. 9, 888245. https://doi.org/10.3389/FNUT.2022.888245.
Yin, J., Hameed, S., Xie, L., Ying, Y., 2021. Non-destructive detection of foreign contaminants in toast bread with near infrared spectroscopy and computer vision techniques. J. Food Meas. Charact. 15 (1), 189–198. https://doi.org/10.1007/s11694-020-00627-6.