

YOLO-Based Deep Learning for Citrus Fruit Detection, Counting, and Yield Estimation in Complex Orchard Environments: A Systematic Review

Paola D'Antonio

University of Basilicata

Luis Alcino Conceição

Polytechnic Institute of Portalegre

Danilo Travascia

University of Basilicata

Lucas Santos Santana

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Josiane Maria da Silva

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Francesco Toscano

`francesco.toscano@unibas.it`

University of Basilicata

Systematic Review

Keywords: YOLO, object detection, citrus fruit detection, yield estimation, precision agriculture, lightweight, systematic review

Posted Date: August 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10773270/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

A systematic review based on 174 Scopus-records of studies using YOLO-type one-stage detectors for detecting citrus fruits, their count, and yield estimation followed PRISMA guidelines 2020. The title/abstract-screening process, done in duplicate ($\kappa=0.920$) yielded 90 included study-records, followed by two further post-hoc exclusions. Each study in the 80 reporting on fruit-level detection showed an average precision of 89.6%, recall of 85.9%, and mAP@0.5 of 91.0%. However, coverage for any individual metric rarely exceeded half of the studies, and only 13% were able to report the more stringent mAP@0.5:0.95. Both YOLOv8 and YOLOv5 were each utilized as the backbone architecture by approximately 22.2% of the studies. From 2025, YOLOv11 has also been emerging. Half of all studies modified architectural components including attention modules, lightweight architectures, and variants of IoU loss functions. Original contributions are generally concentrated in downstream tracking, sensor fusion, and yield modeling rather than the detector itself. A custom-made seven-domain risk of bias tool was developed and utilized by two reviewers who arbitrated discrepancies (91.5%). Results showed that all but one of the reviewed studies had a high level of risk due to almost universal lack of statistical validation and limited dataset diversity; a sensitivity analysis excluding the most risky studies left the performance profiles nearly identical. We conclude that the field has converged around a common technical toolkit but continues to lack standardized benchmarks, multispectral data, and rigorous field-deployment validation.

1. Introduction

The use of machine vision for detecting and counting fruit in place under varying lighting, occlusions, and canopy densities that differ significantly from orchard to orchard and from stage to stage of growth, is critical for yield forecasting, harvest-labor planning and selective robotic harvesting [1, 2]. Citrus is grown in more than 140 countries and is among the most cultivated fruit crops globally. Harvesting manually accounts for approximately 45-60 percent of overall citrus production cost, with throughput as low as 90-120 fruits per worker-hour in commercial operations [3]. Manual counting and estimating yield visually represent the majority of operational defaults in most orchards, however, this scaling model fails to be consistently reliable due to high inter-rater variability. Therefore, the last few decades of machine vision research into automated fruit detection have focused heavily on the development of various approaches using single-stage object detectors such as those within the YOLO (You Only Look Once) family to improve the reliability and consistency of fruit detection.

When we consider the alternatives, such as two-stage detectors like Faster R-CNN and Mask R-CNN, or the older approaches that relied on color and texture segmentation, the YOLO-based models have a distinct advantage because they provide a good balance between detection accuracy and the speed at which they can process images. This balance is the reason why they are well-suited for use on UAVs and on harvesting robots where computational resources may be limited and real-time performance is important. Many researchers have pointed to this balance as the main reason why they chose to use YOLO for their work on fruit detection. Fully transformer-based pipelines (e.g. DETR-style architectures) appear in our corpus solely as occasional comparison baselines rather than as any included study's primary method — for instance [3]'s ACDNet benchmarks against an RT-DETR-R18 baseline and reports it both less accurate and markedly slower than every YOLO variant compared. Another study, [4]'s AG-YOLO, replaces the backbone convolutional network layer of YOLO with a transformer (NextViT) block; we classified it as part of the YOLO family given this structural continuity — it retains the YOLO detection head, neck, and loss framework — and the authors' own naming. Although we recognize AG-YOLO as being partially outside of a mostly convolutional backbones corpus, collectively these findings indicate that the current center of mass in terms of practical applicability for real-time fruit detection remains centred around YOLO-type models for now.

Since YOLO-type models have evolved rapidly since their first release (versions v1 through v11 exist at the time of writing, along with several community maintained versions extending beyond the original Ultralytics lineage), numerous studies have utilized these models to develop solutions to problems related to citrus cultivation. Due to the very large number of studies that utilize these models to address a wide range of problems in citrus cultivation, there exists a highly disjointed body of literature regarding these models. Specifically, this disjointedness manifests in terms of task definitions (i.e., detection, counting, yield prediction), backbone versions, dataset sizes, reported metrics, and hardware targets.

The rapid development of new versions of YOLO has led to extensive reviews of the architecture of successive versions. For example, [5] and [6] provided detailed descriptions of the architecture of successive versions of YOLO (YOLOv1 through v8/v9), including comparisons between successive versions. In addition, [7] expanded upon this previous research by providing a description of the architecture of YOLO11. More generally, within agriculture, [8] surveyed applications utilizing YOLO-type object detection within multiple domains including crop monitoring, livestock monitoring and disease monitoring. More directly applicable to the purpose of this review, [9] recently performed a bibliometric analysis (620 articles) combined with a systematic literature review (114 articles) of YOLO-based fruit detection across all fruit crops. They found a mean detection performance for fruit crops of 91.1%, and found that computational efficiency was reported in only 25.4% of their analyzed studies. Both results demonstrate remarkable convergence given the significantly disparate scopes and methodologies employed by [9] relative to our review. However, based solely on information contained in their abstract, [9] appears to be a broad-scope review that does not focus specifically on citrus, nor do they report on either the PRISMA-style eligibility flow, dual-rater screening reliability or formally assess the risk-of-bias associated with the evidence supporting the reported performance. No previous review, as far as we are aware, has focused on citrus specifically, nor has any previous review applied full-text dual extraction with documented inter-rater agreement, or evaluated the quality of the evidence using a structured risk-of-bias tool; our review aims to supplement this broader work rather than replicate it.

Prior narrative surveys of deep-learning-based approaches for fruit-detection frequently rely on information extracted solely from abstracts and thus grossly under-report quantitative performance. In addressing the above, this review will answer four questions: (RQ1) What is the distribution of reported detection metrics (precision, recall, mAP, F1) among various YOLO-based citrus studies and how does the variability among the distribution differ depending upon the version of YOLO and the specific task? (RQ2) Which architectural modifications were repeated by multiple research groups within the studies reviewed, and can these recurring modifications be identified as a set of techniques that are common to many of the approaches found in the literature? (RQ3) In terms of the datasets used, the hardware platforms employed, and the environments where these systems have been deployed, what are the main characteristics that

define the empirical basis of this research area? (RQ4) When the authors of the primary studies discuss the limitations of their own work, what kinds of issues do they identify, and what do these suggest about the directions that future research in this field should take?

2. Materials and Methods

2.1 Protocol and eligibility criteria

This review followed the reporting guidelines of PRISMA 2020. A prospective definition was made of the review question, inclusion/exclusion criteria, and search strategy; however, there is no record of this being registered separately. Studies included in the review used a model from the YOLO family (including any version thereof), either as the primary method or as part of multiple methods which they compared using baseline performance metrics, that were analyzed on images of fruits from the citrus genus for the purpose of object detection, location identification, count determination, or estimation of yield/production, and provided empirical evaluation(s).

Exclusion reasons:

- (E1) - not concerning citrus at all;
- (E2) - citrus-related but addressing disease, pest, or defect detection rather than the fruit itself.

2.2 Search, screening, and data extraction

Scopus was searched using the following query (last executed mid-2026):

TITLE – ABS – KEY((citrusOR or an ≥ ORCitrus sinensisORm and ar ∈ ORtan ≥ r ∈ eOR ≤ monOR lim eORgrapevit)ANI

We obtained 174 search results. We selected Scopus as the single source of data because it is considered to be the largest collection of literature related to engineering and applied science. It has been reported that approximately 99% of journals contained within the Web of Science (WoS) are also listed in the Scopus database; however, Scopus lists an estimated 21,000 more titles than WoS, including those related to applied science and engineering, along with conference proceedings which are most likely to relate to this subject area. Research comparing the databases of citations have consistently demonstrated that Scopus is virtually a complete subset of WoS in relation to citation of literature in engineering and technology. In addition, Scopus covers a larger number of countries than does the Web of Science [10].

Each title and abstract was independently screened for inclusion/exclusion using the two screening criteria. The agreement among the two reviewers was nearly perfect ($\kappa = 0.920$, per Landis & Koch), where disagreements were resolved through dialogue. The number of articles deemed eligible after screening the titles and abstracts was 92. There were 82 studies that did not meet the eligibility criteria. Of these 82 studies, 76 were excluded because they did not involve citrus (E1), and six were excluded because they involved citrus, but focused on detecting diseases, pests, or defects in the fruit (E2). Additionally, two studies were removed after the authors completed data extraction (Section 2.2). One study [11] was initially included due to its citrus-related index terms; however, upon reviewing the article it became apparent that it actually discussed unmanned boat detection systems for identifying marine surface vehicles instead of citrus. Therefore, this study was considered E1 (i.e., not citrus related) and was applied retroactively. Another study concerned identification of *Diaphorina citri*, the insect responsible for transmitting the Citrus Greening Disease. As such, this study was also considered E2 (i.e., not about the fruit), and therefore was removed. The final number of studies that were included in the review is 90. A diagram illustrating the PRISMA flow of studies can be seen in Figure 1.

The extraction process for the data within the 90 studies consisted of having two independent reviewers extract the same data from the study. Any disagreement that occurred during the extraction of the data was arbitrated through the two reviewers discussing their findings. Prior to beginning to review the entire body of work as it pertained to extraction of the data, both reviewers reviewed a random sample of twenty percent (approximately eighteen studies). They reviewed these studies prior to reviewing the remainder of the studies in order to compare their findings. The results showed perfect agreement between the two reviewers at 100% for all quantitative measures (all discrepancies were less than or equal to 0.5 percentage point difference) when comparing findings for the calibration sample, and also agreed upon task category coding 88.9% (16/18) of the time; however, there were some borderline cases regarding whether the task being completed was "counting" vs. "harvesting," which resulted in some minor disagreements. It was necessary to discuss the matter until a consensus was reached.

After the calibration sample had been successfully used to establish agreement between the two reviewers, they began to extract the data from the full body of work. The two reviewers independently reviewed and extracted data from all 90 studies. Following completion of the extraction process for all studies, there was an overall agreement rate between the two reviewers of 97.2% (106/109 values extracted; all discrepancies ≤ 0.5 percentage points) using quantitative measures and an overall agreement rate of 88.9% (80/90) for task-category coding. These rates are consistent with those reported by the reviewers using the calibration sample. Any remaining disagreements between the reviewers on how to categorize the task being studied were resolved through discussion after referencing the original publication. Extracted fields included: YOLO base version(s); architectural modifications (backbone, neck, attention, loss, head); precision, recall, mAP@0.5, mAP@0.5:0.95, F1-score, and inference speed as reported; dataset size and source; hardware; explicit edge-deployment target; and author-stated limitations.

As a reliability check, extracted figures were cross-checked where possible against independent citations of the same result by other studies in the corpus: the precision reported for three other included studies [4, 12, 13] (90.6%, 84.4%, and 91.0% respectively) and the average mAP@0.5 for a fourth [14] (86.8%,

the mean of two class-specific values reported in that study's own text) are each cited within [15]'s related-work discussion; all four figures matched our independently, separately extracted values to the reported decimal place, supporting the reliability of the extraction process.

2.3 Risk-of-bias assessment

In the absence of a validated risk-of-bias tool specifically designed for object-detection studies, we created our own seven-domain evaluation tool, based off of QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) and PROBAST (Prediction Model Risk Of Bias Assessment Tool). Each study received one of three ratings (Low, Unclear, High) in each of the seven areas. Those areas were: (D1) sample size and diversity; (D2) data-split protocol for training/testing; (D3) whether the results from statistical analyses have been supported by multi-run reporting and/or uncertainty quantification; (D4) completeness of report (whether P, R, and mAP were reported); (D5) verification of hardware and deployment of the models; (D6) how well the performance of the new model compared to previously known baseline models; and (D7) reproducibility of the study (availability of code/data).

We then assigned an overall risk-of-bias rating to each study according to the following ordinal aggregation rule: an overall high-risk rating would be given to a study if at least one of the seven domains had been assigned a high-risk rating; otherwise, an overall unclear-risk rating would be assigned if at least one of the remaining six domains had received an unclear-risk rating; otherwise, an overall low-risk rating would be assigned, since only studies receiving a low-risk rating on all seven domains qualify for such a rating.

Two independent reviewers applied the tool to each of the 90 studies ($90 \times 7 = 630$ total judgments). Inter-reviewer agreement was found to be 91.5% (576/630); disagreements were resolved through arbitration, with the higher-risk rating conservatively selected. "Unclear" was assigned whenever the reviewer could not determine a domain's risk from the information presented in either the full text or the abstract — this occurred most often for domains D2, D6, and D7. Results are summarized in Table 5 (§3.5), together with a sensitivity analysis excluding the highest-risk studies; the study-by-study rating on all seven domains for each of the 90 included studies, not only the aggregated overall rating, is provided in full in the supplementary risk-of-bias CSV accompanying this submission.

2.4 Synthesis approach

Due to significant variation in data, lighting conditions, cultivars, IoU thresholds, and train-test splits among studies, we performed a descriptive quantitative aggregation of means, medians, standard deviations, and ranges, rather than performing a formal meta-analysis that would have combined the effect sizes of studies. Our decision regarding the type of analysis, along with our rationale, is discussed further in §4.1. The architectural changes were coded inductively by two reviewers against an established taxonomy of techniques (attention / backbone / loss / neck / head / downstream-pipeline), while the task category was assigned as a single, mutually exclusive label per study, based upon its primary stated contribution. All statistical analyses were conducted on the 90 included studies wherever possible, using the data reported within the authors' published records.

3. Results

3.1 Study selection, characteristics, and task taxonomy

Fig. 1 summarizes the PRISMA flow. Of 174 records identified, 92 were retained after title/abstract screening; two further post-hoc exclusions during data extraction (§2.2) yielded 90 included studies.

Studies included in this review began in 2020 and have accelerated to the point that 72 percent of the total data examined was collected during the last 3 years of the study period, consistent with a rapidly accelerating research pace rather than a saturating one. The two most productive journals, *Frontiers in Plant Science* ($n = 7$) and *Smart Agricultural Technology* ($n = 6$), are focused on agriculture and technology. They were followed closely by *Agriculture* ($n = 5$); *Computers and Electronics in Agriculture* ($n = 4$); *Agronomy* ($n = 4$); and several other specialized journals focused on the application of technologies to plants. All of these journals represent an "applied" focus, in contrast to the many journals which focus primarily on computer-vision applications. Table 1 describes the two best-performing studies for each of the tasks by mAP@0.5. A complete description of the characteristics of all 90 included studies can be found as Supplementary Table S1.

Table 1. Top two studies by mAP@0.5 within each of the six fruit-level task categories defined in §3.1 (excluding adjacent-scope and benchmark/dataset-only studies, which are not part of this taxonomy; full characteristics for all 90 studies in Supplementary Table S1).

Ref	Task	YOLO base	mAP@0.5 (%)	Precision (%)	Recall (%)
[16]	Robotic harvesting	v5	98.4	—	—
[17]	Robotic harvesting	v5-s	97.7	—	—
[18]	Yield/counting	v5	98.2	—	97.7
[19]	Yield/counting	v11-s	95.5	—	—
[2]	Yield/counting (tracking)	v3	95.7	—	—
[20]	Yield/counting (tracking)	v5-s	94.8	—	—
[21]	Yield/counting (UAV)	v7	86.5	88.5	83.7
[22]	Yield/counting (UAV)	v5mu/v8m/v9c	85.1	—	—
[23]	Maturity classification	v5	94.2	93.2	89.6
[24]	Maturity classification	v8	93.4	88.6	93.1
[25]	Fine localization	v8-n	94.4	96.0	90.9
[26]	Fine localization	v5-s	88.5	92.3	79.1

YOLOv8 and YOLOv5 were tied as the dominant base architectures (20/90, 22.2% each), together accounting for 44.4% of the corpus; YOLOv4 followed (12/90, 13.3%), YOLOv7 and YOLOv11 (10/90, 11.1% each, the latter concentrated in 2025–2026, consistent with its late-2024 release by Ultralytics), and YOLOv3 (8/90, 8.9%). Fig. 2a plots the version distribution against publication year.

We distinguish six task categories beyond simple "detection", assigned as a single, mutually-exclusive label per study reflecting its primary stated contribution (inter-reviewer agreement across the full corpus: 88.9%, §2.2): static image-based **yield estimation/counting** (26/90, 28.9% — [1, 12, 18, 19, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]); **video- or tracking-based counting** (5/90, 5.6% — [2, 20, 49, 50, 51]); **UAV-based** yield estimation (5/90, 5.6% — [21, 22, 52, 53, 54]); **harvesting-robotics-oriented detection** (28/90, 31.1%, the largest single category — [3, 4, 13, 15, 16, 17, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]); **maturity/ripeness classification** (5/90, 5.6% — [23, 24, 77, 78, 79]); and **fine localization** (fruit pose, diameter, volume, or mass estimation; 4/90, 4.4% — [25, 26, 80, 81]). A further 10/90 (11.1%) studies concern citrus but not fruit-level detection — canopy/trunk detection [82, 83, 84], orchard 3D reconstruction [85], post-harvest grading [86], phenological organ detection [14], cargo-logistics robotics [87], a secondary generalization test [88], a tactile-sensing platform [89], and an AI-enabled orchard monitoring system [90]. We report these ten as "adjacent-scope" throughout; §4.1 discusses scope sensitivity. A further 6/90 (6.7%) studies were pure benchmark/dataset comparisons without a task-specific focus of their own [91, 92, 93, 94, 95, 96]. The single remaining study [97] had insufficient detail in its published record for confident task classification.

3.2 Performance synthesis and architectural modification taxonomy

Table 2 reports the descriptive synthesis of detection performance across the fruit-level corpus.

Table 2. Descriptive synthesis of reported detection performance across fruit-level studies (n=80).

Metric	N	Mean	Median	SD	Range
Precision (%)	28	89.6	89.9	4.2	78.2–96.0
Recall (%)	25	85.9	85.4	7.6	69.7–100.0
mAP@0.5 (%)	40	91.0	92.6	5.9	75.0–98.4
mAP@0.5:0.95 (%)	10	66.2	66.3	14.4	46.3–87.1
F1-score (%)	8	92.6	92.5	3.5	86.0–98.0

Note: Not all studies reported all metrics; sample sizes therefore vary by metric. The 10 adjacent-scope studies described in §3.1 were excluded. Estimates for F1-score (n=8) and mAP@0.5:0.95 (n=10) are illustrative because of the limited number of reporting studies.

The coverage itself is a clue as to why this may be so; mAP@0.5 was extractable for 40/80 fruit-level studies (50.0%) while the COCO-style mAP@0.5:0.95 was extractable for only 10/80 (12.5%), clearly suggesting that much of the field continues to report against the less strict IoU=0.5 convention. Recall (69.7–100%) was consistently the noisier metric compared to Precision (78.2–96.0%), an observation consistent with occlusion — which is far more destructive of recall through its generation of false negatives on hidden fruit, as opposed to false positives — being the field's central, unaddressed issue (§3.4).

Ranked by mAP@0.5, the highest-performing studies were [16] (98.40%), [18]'s YOLOv5-CS (98.23%, with recall 97.66% and demonstrated edge deployment), [17]'s BCo-YOLOv5 (97.70%, though evaluated across citrus, apple, and grape jointly), [3]'s ACDNet (97.50%, with precision 92.1% and recall 92.8%), [55]'s Citrus-YOLOv7 (97.29%), and [56]'s FAP-YOLOv8 (97.20%). By mAP@0.5:0.95, [3]'s ACDNet is the strongest performer (87.1%). The study also reported variance in outcomes between different runs, in addition to [27]. It had the greatest amount of variance reported between runs — averaging ten independent

runs for each condition, and calculating Cohen's d as well as paired t-tests. In assessing the risk of bias (§3.5), [3] rated Low for D1 (dataset), D3 (statistical methods used), and D4 (completeness of reporting). However, [3] did not rate the study as having a low risk of bias overall. Rather, it was rated as having an unclear risk of bias, due to several factors – lack of information about how transparent the protocol was, whether deployment was validated, what type of baseline comparison was made, and whether the results could be reproduced (D2, D5, D6, D7). This is flagged here so that the study's evident statistical rigour on D3 is not read as a blanket endorsement across all seven domains. Table 3 summarizes the architectural and pipeline modifications identified across the corpus.

Table 3. Architectural and pipeline modification taxonomy across the 90 included studies.

Category	Studies (n/90)	Most-used techniques (n)
Attention mechanism	25	CBAM (7), BiFormer (5), Transformer (5), SE (3), illumination-adaptive attention (1), other (9)
Lightweight backbone	18	FasterNet/PConv (6), GhostNet (4), MobileNet (3), pruning/distillation (3), DSConv (3)
Extra detection head/layer	11	Small-object (P2/P6) layer (8), anchor re-clustering (2), dynamic multi-scale sampling (1)
Neck/FPN modification	10	BiFPN (3), GSConv/VoVGSCSP (3), PMCE (1), AFPN (1), CGFPN (1), Bi-PANet (1)
IoU-loss variant	10	SIoU, EIoU, WIoU, MPDIoU, Shape-IoU, PIoU, FSA-IoU (1–2 studies each)
Downstream: multi-frame tracking	7	SORT/DeepSORT-type (5), Kalman + Hungarian (2)
Downstream: sensor/model fusion	6	Depth/point-cloud (2), ensemble ML regression (2), geometric volume/mass modelling (2)
None of the above	44	Direct application, pure benchmark, dataset-only, input-preprocessing ablation, or insufficient architectural detail in the available record

Attention mechanisms were the single most common modification category (25/90 studies): CBAM (Convolutional Block Attention Module) alone accounted for 7, followed by BiFormer (5) and self-attention/Transformer components including illumination-adaptive variants (e.g. [3]'s CAFE module, [29]'s MSDA). Lightweight-backbone strategies (18/90) were led by FasterNet/PConv (6) and GhostNet (4). An additional small-object detection layer appeared in 11 studies, and an IoU-loss variant replacing the default CloU (Complete Intersection-over-Union) loss in 10.

The two other categories that capture innovations in the processing of data "downstream" of a detector are multi-frame tracking techniques designed to avoid double-counting (7 studies, most rigorously in [2], where multi-frame tracking reduced the mean absolute error (MAE) in counting from 0.45–1.21 to 0.081), and integration with additional sensor(s) or geometric yield model(s) (6 studies, including [29]'s ellipsoid-volume mass estimation, validated with $R^2=0.995$). The catalogued techniques were used by forty-six of ninety studies (51.1%); the remainder (forty-four of ninety, or 48.9%) consisted of direct applications, benchmark studies, dataset-only studies, and/or those studies where there was insufficient documentation of their architecture in the published literature to classify them.

Performance on tasks has systematic variation, as shown in Table 4: the highest mean mAP@0.5 for robotic-harvesting-oriented detection studies was reported (n = 14) at 93.1%, while the lowest mean mAP@0.5 for UAV-based yield estimation was reported (n = 2) at 85.8%. We found no correlation between study size and reported mAP@0.5 (Pearson $r = 0.06$, $p = 0.76$; Spearman $\rho = 0.16$, $p = 0.44$; n = 26 studies reporting both). This null result remained when we transformed study size to logarithmic values ($r = 0.24$, $p = 0.23$), which matches what would be expected if between-study variations in reported accuracy are due to differences in the difficulty of the task(s) being evaluated and/or the conditions under which the evaluations are performed, rather than due to differences in the amount of data available. However, since the study sample is relatively small and non-randomized, this finding should not be relied upon too heavily. We recognize that the data set is not composed of entirely independent observations as defined by classical statistics – a relatively few numbers of author groups include more than one study. Thus, we did not adjust for this using a mixed effects model with an authors' group random intercept; since n = 26, and at most 2-3 studies/author group, such a model would be very poorly identified. Therefore, we report on the relationship here with caution, and do not provide an overly complicated analysis where our measure of precision may be misleading.

Table 4. Reported mAP@0.5 by task category (studies reporting mAP@0.5 only; fruit-level detection tasks; adjacent-scope studies excluded per §3.1).

Task category	N	Mean	Median	SD
Robotic harvesting detection	14	93.1	94.7	5.6
Yield/counting (tracking)	4	90.9	93.8	7.2
Yield/counting (static)	11	89.7	92.1	7.2
Maturity classification	4	91.5	93.1	3.8
Fine localisation	2	91.5	91.5	4.2
Yield/counting (UAV)	2	85.8	85.8	1.0
Benchmark/dataset studies	3	88.1	89.2	3.7

12/90 (13.3%) studies explicitly targeted edge hardware. Inference speed was reported by 23/90 studies (25.6%), spanning from 168 FPS (frames per second) on a 0.81M-parameter model [28] to well under 30 FPS for larger backbones — a range too hardware-dependent to summarise as a single pooled figure

3.3 Dataset and hardware characteristics

The number of images used in each study was reported for 46 out of 90 studies (51.1%), which was found to be a highly positively skewed distribution (average = 2365 vs. median = 1485 images; range = 100-19320). Most studies are based upon small datasets from a single location; only CitDet, an open-source dataset that is used by [27], as well as probably at least one other pair of shared datasets ([91, 92]; both N=5025) were found. Only two studies utilized multispectral data [30, 93], therefore there appears to be a significant void in this area. Of the hardware platforms that were described, the Jetson family of GPU-based edge devices were the most common platforms utilized in studies (n=11), followed by a variety of desktop/server GPU-based systems and then Raspberry Pi CPU-based systems (n=2 studies).

3.4 Persistent limitations (author-reported)

A qualitative review of the limitation and future-work sections — a limited "pass" flagging notable themes that were mentioned explicitly in each study, rather than an exhaustive frequency count — identified six recurrent themes. The problem of occlusion was a major recurring issue, with studies reporting failure beyond certain thresholds (e.g., above 50% foliage occlusion in [13]); given that we did not conduct an exhaustive search across all studies' discussions, we cannot report that it was the single most frequent theme, however, it was the one most often accompanied by a specific, quantified failure threshold. Limited dataset diversity was acknowledged in several studies [52, 53, 77, 84, 86, 91, 93]. Green-on-green camouflage was identified as a hard sub-case in three studies [15, 27, 57]. A laboratory-to-field performance gap was directly quantified by [58] (recall 79.8% → 75.1% on a deployed robot). Several studies identified missing 3D/spatial information [2, 59], and one noted network-connectivity dependency for multi-UAV systems [90]. [3] and [92] both call for validation under extreme environmental conditions. We did not systematically tally how many of the 90 studies mention each theme, since limitations sections are reported with very uneven detail across the corpus; the counts given above are lower bounds reflecting only the studies where a theme was salient enough to record during extraction, not a complete census.

3.5 Risk-of-bias assessment

Table 5 summarises the risk-of-bias ratings across all seven domains for the full corpus.

Table 5. Risk-of-bias summary across all 90 included studies (seven-domain assessment, two independent reviewers with arbitration; inter-reviewer agreement 91.5%).

Domain	Low	Unclear	High
D1: Dataset size & diversity	18	19	53
D2: Protocol transparency	13	67	10
D3: Statistical validation	2	1	87
D4: Reporting completeness	19	30	41
D5: Deployment validation	12	28	50
D6: Baseline comparison	40	44	6
D7: Reproducibility	1	81	8
Overall	0	1 (1.1%)	89 (98.9%)

A larger number of studies are at high risk due to the primary source of high risk being D3 (statistical validation). Of the 90 studies in this analysis, 87 (or 96.7%) were reported to have been run once through training/evaluation with no quantified measure of uncertainty. Studies [3] and [27] were the only ones that provided information regarding multiple runs of their testing. The second most common area for a study to be identified as high risk was dataset size/diversity (D1), which was present in 53 of the studies analyzed (or 58.9%) — in other words, these studies either contained fewer than 500 images or failed to provide an image count. No study achieved Low risk on every domain when evaluated using the method outlined in §2.3, while the sole study that received an overall Unclear risk assessment failed to achieve a High rating on every domain, but also did not receive a sufficient number of Low ratings to be assessed as overall Low. This study is [3] (§4.1), which is Low — not High — on D1 and D3, and Unclear (not High) on D2, D5, D6, and D7; its overall rating is Unclear because Unclear ratings appear on four domains, none of which reaches High, illustrating that the aggregation rule is driven by the single worst rating across domains, not by a count or average of risk levels. Because our primary rule ("High if any domain is High") is deliberately conservative, we also computed two alternative aggregations directly from the same seven domain-level ratings, without any new extraction: an "any-High" ultra-conservative check (§3.5, below) and a less severe "≥2 domains High" rule, under which 83/90 studies (92.2%) — rather than 89/90 (98.9%) — would be classified overall High. That the two rules converge on a similarly high proportion indicates the finding is not an artefact of the specific threshold chosen; the full per-domain ratings (Table 5) remain available for readers who prefer to apply a different aggregation rule, or none at all, to our seven-domain data (Supplementary risk-of-bias CSV).

Sensitivity analysis. We focus the primary sensitivity check on D1 (dataset) and D3 (statistics) specifically, rather than on the overall rating or all seven domains jointly, because these two domains bear most directly on whether an individual study's *reported accuracy figure itself* can be trusted: an unreported or very small dataset limits how representative the accuracy estimate is, and the absence of statistical validation means the reported figure is a single point estimate with unknown variance. The remaining five domains (D2, D4–D7) chiefly concern reporting transparency, deployment realism, and reproducibility, which affect the trustworthiness of the *study as a whole* but not the arithmetic validity of the specific number entering Table 2's pooling. Excluding the 51 studies with concurrent high risk on both D1 and D3 — the most methodologically vulnerable subgroup for this purpose — and recalculating performance on the remaining fruit-level studies (35 of the 80 addressing fruit-level detection, §3.1) yielded: mAP@0.5 mean 90.7% (median 92.3%, n=20), precision mean 88.0% (n=11), recall mean 84.0% (n=11), closely matching the fruit-level corpus estimates (Table 2). As a stricter, more conservative alternative, we also examined performance after excluding every study with *any* High-risk domain rather than only D1+D3: this leaves exactly 1 fruit-level study ([3], mAP@0.5 = 97.5%) with no High rating on any of the seven domains, precluding any meaningful mean or sensitivity comparison. This is itself an informative finding — it shows that the corpus contains virtually no studies free of high risk on at least one domain, and that a fully conservative filter is too strict to support quantitative synthesis at all with the present evidence base. We therefore report the D1+D3 criterion as the primary, most defensible sensitivity check, while noting this more extreme alternative explicitly rather than omitting it.

4. Discussion

4.1 Principal findings

The area of YOLO-based citrus detection appears to be an active and relatively successful technical space, based on the cited figures; however, this success occurs within a fragmented and poorly developed methodology. The risk-of-bias assessment (§3.5) indicates that approximately 96.7% of studies examined in our analysis failed to include any type of statistically validated multiple-run evaluation (D3), while approximately 58.9% of all studies evaluated were either limited to fewer than 500 images or did not specify the total number of images utilized (D1). While authors have reported median values for precision and mAP@0.5 greater than ninety percent (Table 2), this apparent convergence in reported median values masks the extreme variability observed among different authors regarding the manner in which outcome measures are reported, upon what type of data, and under what conditions the architectures are deployed — and no author weights their results according to a measure of quality-of-evidence.

Similarly, the general architecture of the detectors (§3.2) supports our finding of great variability: although a set of techniques (attention modules, lightweight backbone networks, various forms of IoU loss) is commonly combined across a large number of independent groups, the variety of ways in which these combinations occur does little to suggest that a majority of the research conducted within this area represents anything other than convergence on a small number of effective component parts. In fact, it is our view that a larger proportion of the more novel contributions made by authors can be located downstream of the detector itself, in areas such as object tracking, sensor fusion, and geometric yield/mass modeling — and we include our reservations regarding this assertion.

This downstream trend is important because the ultimate benefit of these detection systems will depend on their ability to integrate into reliable yield-prediction systems. Recent bibliometric analyses examining related Mediterranean tree crops have shown that, despite advances in several areas (sensor fusion, open data infrastructure, explainability of artificial intelligence models), large gaps remain in each of these areas, which impede widespread adoption of precision-agriculture forecasting [98]. As such, we provide descriptive statistics for the studies included in our analysis, and we do not perform a formal meta-analysis. Performing a meta-analysis of the mAP results from studies that differ significantly in terms of crop variety, lighting, labeling protocols, IoU thresholds, and training/testing splits would produce a numerical summary, but one that could not accurately be described as a meaningful measure. Furthermore, we recognize that many pooled mAP values are not directly comparable. Most studies report single-class fruit mAP values, while others use average mAP values calculated using multiple maturity classification outputs.

Finally, due to the fact that no shared benchmark dataset has been utilized by more than two studies, it follows that we cannot pool mAP values through meta-analysis. Similarly, we also cannot calculate traditional measures of heterogeneity (Cochran's Q and I^2). Both Q and I^2 require that each study report a standard error or confidence interval for the pooled metric, and subsequently, both require that the variance estimate for the pooled metric be derived from repeated runs or cross-validation estimates — both of which types of variance estimates occur in only 2 of the 90 studies examined in §3.5. Therefore, we do not report I^2 .

There are two boundary decisions that may warrant further explanation. The first involves the removal from consideration of one document (§2.2), which illustrates a potential failure mode of utilizing databases in screening for inclusion — one that can only be caught through title/abstract review.

Additionally, the ten studies located immediately adjacent to the scope boundaries (§3.1) — while characterized similarly to those studies within the boundaries — were removed from consideration during performance analysis (see Table 2), representing a reasonable scope judgment. Studies focused on canopy- or trunk-level phenomena are clearly operationally relevant to estimating yields, even though they fall outside fruit-level detection, as canopy volume and structure have been identified as indirect-yield predictors. To confirm this, we also re-ran the calculation for Table 2 including the additional performance metrics provided for the adjacent-scope studies — and each mean shifted by less than 1 percentage point. In other words, the scope decision has no material impact on the overall central tendencies presented in Table 2; it is intended solely to clarify what population "detection performance" should refer to, not to change the actual values.

Finally, the limitations synthesis (§3.4) identifies occlusion as another area with ongoing issues — particularly since occlusion-handling modules are also among the most commonly implemented architectural interventions (§3.2). This finding can be viewed as a measure of the degree to which the body of literature demonstrates relative improvement over study-specific baseline levels, compared to absolute levels achieved under standardized severe-occlusion

conditions — i.e., a measurement gap rather than necessarily a methods gap. The laboratory-to-field degradation demonstrated by [58] suggests that there are real-world operational implications associated with this measurement gap; however, fewer than half of all studies evaluate their model(s) using deployed hardware.

4.2 Implications for future research

There are three priorities which arise out of the areas where there is still significant need. Firstly, very little research has been done on using either multispectral or hyperspectral data to assess crop health (only 2 of the 90 studies reviewed). We recognize that the use of multispectral sensors can be expensive and difficult to integrate onto platforms suitable for smallholder applications; researchers therefore need to consider whether increases in accuracy can be realized versus improved deployability. Secondly, development of common benchmark datasets would allow reviewers to pool their results, and potentially limit duplication of effort. The third aspect is that future research may also have diminishing returns unless accompanied by the scientifically valid, field-based validation efforts demonstrated by [3] and [60][94]. Lastly, a fourth area of importance is what we refer to as "open science," which is further emphasized by the risk-of-bias evaluation (§3.5). Only one study in the body of literature evaluated utilized a publicly available dataset, and no studies referenced using open-source coding methods — both of which impede the ability of other researchers to reproduce studies and compare them across different studies.

4.3 Limitations of this review

We have identified 5 possible limitations of this systematic review. Firstly, the search was limited to Scopus. A number of comparative analyses demonstrate that Scopus indexes around 99% of journals in the Web of Science [10] plus ~21,000 extra titles, especially good coverage of engineering and conference proceedings. In order to provide an additional check of this limitation, we searched IEEE Xplore — a major source of engineering conference literature — using an analogous search string directly in IEEE Xplore, with results returning 77 records. After manually screening these based on the eligibility criteria, we found that 12 of the studies had already been retrieved in our initial search from Scopus; 3 of the studies reported in IEEE Xplore were excluded from our corpus due to meeting the E2 criterion (pest and psyllid detection, instead of fruit detection), exactly the type of inclusion errors our subsequent exclusion process was intended to catch (§2.2); the majority of the rest were false positives that occurred solely due to coincidence in the occurrence of query terms (i.e., "orange" as a color/hardware description, "lime" as an unrelated LIME explainability method); and 3 additional studies were identified as being citrus eligible and missing from our final included corpus.

In order to determine whether a genuine gap existed within the Scopus database for these 3 missing studies, we conducted a separate Scopus search for each of them. Of those, two — a lemon-detection study and a lemon-ripeness-detection study — were confirmed to actually be missing from Scopus, both in a new search that was conducted after this review's original data extraction, confirming there was indeed a relatively minor database-coverage gap. The third record, an orange/lemon-detection-and-image-augmentation study, was found to exist in our initial 174-record export of results from Scopus and exists still. However, the third study did not meet our inclusion criteria (E1 or E2) and therefore was not selected for inclusion. Unfortunately, we were unable to determine why it was rejected from our corpus during our manual screening of the studies' titles and abstracts.

Therefore, as opposed to retroactively creating a reason for exclusion, we are reporting it transparently as an apparent single-study screening discrepancy; it serves as an example of how some level of subjective decision making applies not just to extracting data from full texts (§4.3, "Third"), but also can apply to screening. The process described above is more persuasive than merely debating whether the databases provide adequate coverage on this topic, to establish that no reviewer decision made at any time throughout this review can ever be deemed absolute. We did not include either of the two studies that were based solely on IEEE, or the study that was mis-screened in Scopus, in our final included body of evidence for purposes of re-running extraction and conducting a new risk-of-bias assessment, because selectively choosing only three records to re-screen from a much larger group of records, without also screening the remainder of the corpus, would introduce yet another form of bias; we will note them for both transparency and for possible revisions or updates to this review; full bibliographic details for all three IEEE-identified records are given in Supplementary Table S2 rather than in the main reference list, since none was included in our corpus. In order to provide an additional, independent measure of the extent of database coverage, we used the exact same search string, translated this time into Web of Science Core Collection format, which resulted in 128 unique references. The majority of these references were either already present in our corpus, or clearly did not belong to the citrus domain based solely upon the titles of those references (i.e., unrelated areas such as fish, poultry, medical imaging, or infrastructure inspections). These records were identified as false positives where the query terms merely occurred coincidentally within those reference titles, similar to how the IEEE Xplore false positives occurred earlier. There were also a small number of records related to citrus but still ineligible for being "citrus-adjacent" (e.g., pest or psyllid detection, meeting our E2 criterion). We will make some mention of two records individually; full citations for both can be found in Supplementary Table S2. One genuine citrus fruit-level ripeness-detection study, using a modified version of YOLOv8 architecture, was not located in either of the two Scopus export datasets generated, providing another example of the coverage gap previously noted. A second record proposed a custom one-stage detector specifically for localization of the stem-end and blossom-end of navel oranges. This second record existed in Scopus, but was not returned by our search string. On inspection, this detector was not itself developed around YOLO — YOLOv2 and YOLOv3 are presented as just two of six different baseline architectures the authors benchmark their own model against, rather than as the study's primary method. Therefore, this would be a borderline case that would most likely not have met our primary-method eligibility criterion (§2.1), even had our search string captured it. Neither record was added to the included corpus, for the same reasons of selection consistency given above.

Secondly, not all studies report all metrics; the lowest reported metric is 8/80 for F1-score for fruit-level studies (the smallest sample in Table 2), whose mean and standard deviation should therefore be taken as illustrative rather than reliable — and this under-reporting is another significant finding regarding reporting practices in this field — although it reduces our ability to synthesize performance results precisely.

Thirdly, while two independent reviewers extracted data across the entire corpus with relatively high levels of agreement (§2.2), there still remained a degree of disagreement when classifying tasks into categories (88.9%) and a few numerical values (97.2%), each resolved through discussion/arbitration; despite the fact that the judgement calls in borderline cases contain some indelible subjectivity, we have presented raw percent agreement rather than a kappa statistic corrected for chance (e.g. Cohen's κ) for classification of tasks into categories specifically, since our calibration sample (20%, $n = 18$) is insufficiently large to yield a stable multi-category kappa estimation — whereas, in contrast, kappa (0.920) was calculated across the entire set of screening titles/abstracts ($n = 174$).

Fourthly, there was no pre-protocol registration of any proposed method (for example through PROSPERO, the International Prospective Register of Systematic Reviews); the risk-of-bias assessment tool (§2.3) that was employed to conduct this review was created specifically for this review and has not been independently validated, but does have good inter-reviewer agreement (91.5%); and can be scrutinized and replicated by others because it includes a clear rule for aggregating ratings across domains — an overall rating of "High" if one or more domains is rated "High," "Unclear" if at least one of the remaining domains is rated "Unclear," and "Low" only when all seven domains are rated "Low."

Fifthly, we did not collect citrus cultivar (sweet orange, mandarin, lemon, pomelo) systematically as an organized field throughout our data collection; citrus cultivar potentially confounded measured performance because of large differences in fruit color, size, and canopy structure among cultivars, which can impact the ease or difficulty of detecting a fruit (e.g., green-on-green camouflage is primarily dependent upon both the citrus cultivar and the stage of ripeness, see §3.4); however, due to significant variability in reporting and describing whether and how citrus cultivar was described in the published literature that existed at the time we collected the data, we cannot accurately extract this information in an organized manner post hoc. A cultivar-stratified re-analysis is therefore a natural extension for future work.

5. Conclusion

Across 90 studies, with 80 focusing on detecting, counting, or estimating yields for individual fruits and 10 focusing on other adjacent citrus-orchard related tasks (§3.1), the central-tendency performance reported by YOLO-based citrus fruit detection was highly consistent across the fruit-level studies (Median $mAP@0.5 = 92.6\%$). Whether the consistency represents an adequate engineering validation of performance or there remains headroom in performance for the models applied that has yet to be explored is a question our descriptive synthesis does not enable us to answer. A purpose-designed risk of bias assessment determined that greater than 90 percent of the studies were lacking in the application of multi-run statistical testing. However, the results from a sensitivity analysis indicated that this lack of multi-run statistical testing would have no material impact upon the distribution of performance measures. The field's most pressing gaps are not in detector accuracy but in measurement standardization, dataset diversity, shared benchmarks, and field-deployment validation.

Declarations

Data Availability

The extracted study-level dataset, the risk-of-bias assessment, and the PRISMA screening records accompany this submission as supplementary materials.

Funding

This work was supported by the University of Basilicata; "Casa delle Tecnologie Emergenti" of Matera, "Laboratorio del Giardino delle tecnologie emergenti" (CTEMT); the "Territorial Pacts" (Patti Territoriali) doctoral research project of F.T. (CUP: C42C24000220001); the Smart@Irrifert Project (PSR Basilicata 2014–2020–Sottomisura 16.2, CUP: G19J21004870006); and the "Vertical gaRdens with Digital twin of plants" (Viridiis) Project, PRIN 2022 (CUP: C53D23000480006)

Author Contributions

Paola D'Antonio: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Francesco Toscano: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Luis Alcino Conceição: Conceptualization, Investigation, Writing – review & editing, Validation.

Danilo Travascia: Conceptualization, Investigation, Data curation, Writing – review & editing.

Lucas Santos Santana: Conceptualization, Investigation, Data curation, Writing – review & editing, Validation.

Josiane Maria da Silva: Investigation, Data curation, Writing – review & editing.

All authors have read and agreed to the published version of the manuscript.

Ethics Declarations

Not applicable. This study is a systematic review of previously published literature and did not involve human participants, animal subjects, or the collection of primary data.

Acknowledgments

This work was carried out within the framework of research activities and scientific collaboration between the University of Basilicata and international institutions, including the Federal University of Vale do Jequitinhonha e Mucuri (Brazil) and the VALORIZA—Research Center for Endogenous Resource Valorization at the Polytechnic Institute of Portalegre, Portugal. This research falls within the scientific-disciplinary fields of Agricultural Engineering.

AI Declaration

Google Gemini 3.1 Pro (Google LLC, Mountain View, CA, USA) was used during manuscript preparation for language editing, writing refinement. The tool was not involved in data collection, statistical analysis, hypothesis development, or interpretation of results. The authors revised all AI-assisted text and bear full responsibility for the published content.

Competing Interests

The authors declare no competing interests.

References

1. Mirhaji, H., Soleymani, M., Asakereh, A., Abdanan Mehdizadeh, S. (2021). Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2021.106533>
2. Zhang, W., Wang, J., Liu, Y., Chen, K., Li, H., Duan, Y. et al. (2022). Deep-learning-based in-field citrus fruit detection and tracking. *Horticulture Research*. <https://doi.org/10.1093/hr/uhac003>
3. Wang, Z., Xia, W., Li, M. (2026). ACDNet: Adaptive Citrus Detection Network Based on Improved YOLOv8 for Robotic Harvesting. *Agriculture*. <https://doi.org/10.3390/agriculture16020148>
4. Lin, Y., Huang, Z., Liang, Y., Liu, Y., Jiang, W. (2024). AG-YOLO: A Rapid Citrus Fruit Detection Algorithm with Global Context Fusion. *Agriculture (Switzerland)*. <https://doi.org/10.3390/agriculture14010114>
5. 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. *Machine Learning and Knowledge Extraction*, 5(4), 1680-1716. <https://doi.org/10.3390/make5040083>
6. Hussain, M. (2023). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines*, 11(7). <https://doi.org/10.3390/machines11070677>
7. Ali, M.L., Zhang, Z. (2024). The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection. *Computers*, 13(12). <https://doi.org/10.3390/computers13120336>
8. Kamalesh Kanna, S., Kumaraperumal, R., Pazhanivelan, P., Jagadeeswaran, R., Prabu, P.C. (2024). YOLO deep learning algorithm for object detection in agriculture: a review. *Journal of Agricultural Engineering*, 55(4). <https://doi.org/10.4081/jae.2024.1641>
9. Pugazhendhi, P., Badgujar, C.M., Sapkota, R., Dhillon, R., S. R., J. J.J.S., Ganapathy, M.R. (2026). Advances in agricultural fruit detection using you only look once (YOLO) algorithm: A review. *Smart Agricultural Technology*, 13. <https://doi.org/10.1016/j.atech.2026.101896>
10. Singh, V.K., Singh, P., Karmakar, M., Leta, J., Mayr, P. (2021). The Journal Coverage of Web of Science, Scopus and Dimensions: A Comparative Analysis. *Scientometrics*, 126(6), 5113-5142. <https://doi.org/10.1007/s11192-021-03948-5>
11. Zheng, S., Yu, S., Ai, X., Peng, P., Liang, J. (2025). Research on Lightweight YOLOv5s Applied in Unmanned Boats. *IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC*. <https://doi.org/10.1109/IMCEC66174.2025.11332005>
12. Liao, Y., Li, L., Xiao, H., Xu, F., Shan, B., Yin, H. (2025). YOLO-MECD: Citrus Detection Algorithm Based on YOLOv11. *Agronomy*. <https://doi.org/10.3390/agronomy15030687>
13. Deng, F., Chen, J., Fu, L., Zhong, J., Qiao, W., Luo, J. et al. (2024). Real-time citrus variety detection in orchards based on complex scenarios of improved YOLOv7. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2024.1381694>
14. Liu, W., Jian, C., Zhang, S., Cong, G., Wang, H., Li, J. et al. (2025). Enhanced YOLOv5s with BiFormer attention for citrus spring shoot detection: Optimized phenological period and cross-regional application. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2025.101446>
15. Yang, J., Wu, H., Lv, D., Ma, W., Teng, H., Chen, D. (2026). A Lightweight and High-Precision Citrus Detection Model for Unstructured Orchard Environments. *Horticulturae*. <https://doi.org/10.3390/horticulturae12060718>
16. Liu, X., Li, G., Chen, W., Liu, B., Chen, M., Lu, S. (2022). Detection of Dense Citrus Fruits by Combining Coordinated Attention and Cross-Scale Connection with Weighted Feature Fusion. *Applied Sciences (Switzerland)*. <https://doi.org/10.3390/app12136600>
17. Yang, R., Hu, Y., Yao, Y., Gao, M., Liu, R. (2022). Fruit Target Detection Based on BCo-YOLOv5 Model. *Mobile Information Systems*. <https://doi.org/10.1155/2022/8457173>
18. Lyu, S., Li, R., Zhao, Y., Li, Z., Fan, R., Liu, S. (2022). Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System. *Sensors*. <https://doi.org/10.3390/s22020576>
19. Li, J., Miao, Y., Guo, W., Li, Y., Xu, Q., Xiang, Y. et al. (2026). High-Throughput Citrus Detection via Citrus-SGYOLOv2: A Symmetric Ghost-Based Architecture with High-Resolution Feature Fusion. *Agronomy*. <https://doi.org/10.3390/agronomy16090894>
20. Lin, Y., Hu, W., Zheng, Z., Xiong, J. (2023). Citrus Identification and Counting Algorithm Based on Improved YOLOv5s and DeepSort. *Agronomy*. <https://doi.org/10.3390/agronomy13071674>

21. Subeesh, A., Prakash Kumar, S., Kumar Chakraborty, S., Upendar, K., Singh Chandel, N., Jat, D. et al. (2024). UAV imagery coupled deep learning approach for the development of an adaptive in-house web-based application for yield estimation in citrus orchard. *Measurement: Journal of the International Measurement Confederation*. <https://doi.org/10.1016/j.measurement.2024.114786>
22. Tanimoto, Y., Zhang, Z., Yoshida, S. (2024). Object Detection for Yellow Maturing Citrus Fruits from Constrained or Biased UAV Images: Performance Comparison of Various Versions of YOLO Models. *AgriEngineering*. <https://doi.org/10.3390/agriengineering6040243>
23. Zhou, B., Wu, K., Chen, M. (2024). Detection of Gannan Navel Orange Ripeness in Natural Environment Based on YOLOv5-NMM. *Agronomy*. <https://doi.org/10.3390/agronomy14050910>
24. Deng, F., He, Z., Fu, L., Chen, J., Li, N., Chen, W. et al. (2025). A new maturity recognition algorithm for Xinhui citrus based on improved YOLOv8. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2025.1472230>
25. Li, H., Yin, Z., Zuo, Z., Pan, L., Zhang, J. (2025). Precision citrus segmentation and stem picking point localization using improved YOLOv8n-seg algorithm. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2025.1655093>
26. Zhang, G., Li, L., Zhang, Y., Liang, J., Chun, C. (2025). Citrus pose estimation under complex orchard environment for robotic harvesting. *European Journal of Agronomy*. <https://doi.org/10.1016/j.eja.2024.127418>
27. Rahayu, S., Andono, P.N., Pujiono, P., Fanani, A.Z. (2026). Evaluation of Color, Texture, and Structural Representations for Green Citrus Detection under Green-on-Green Camouflage. *International Journal of Intelligent Engineering and Systems*. <https://doi.org/10.22266/ijies2026.0831.34>
28. Zeng, T., Tong, J., Sun, X., Liu, J., He, X., Guan, Z. et al. (2025). YOLO-FCAP: An improved lightweight object detection model based on YOLOv8n for citrus yield prediction in complex environments. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2025.101590>
29. Zhang, D., Chu, H. (2026). Citrus instance segmentation and yield prediction based on parallel multiscale channel enhanced. *Scientia Horticulturae*. <https://doi.org/10.1016/j.scienta.2026.114895>
30. Oviedo Espinosa, M.R., Porto, L.R., Orlando, V.S., Tommaselli, A.M., Dal Poz, A.P., Imai, N.N. (2024). Evaluation of YOLO Efficiency in Automatic Orange Detection in Multi-Exposure Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprs-annals-X-3-2024-303-2024>
31. Adeem, G., Abdulkader, O., Ikram, M.J., Aqib, M., Hafeez, Y., Tahir, M.N. et al. (2026). Efficient deep learning-based approach for detecting citrus fruits. *International Journal of Agricultural and Biological Engineering*. <https://doi.org/10.25165/ijabe.20261903.9702>
32. Al Riza, D.F., Musahada, L.C., Aufa, R.I., Hermanto, M.B., Nugroho, H., Hendrawan, Y. (2025). Comparative study of citrus fruits (Citrus reticulata Blanco cv. Batu 55) detection and counting with single and double labels based on convolutional neural network using YOLOv7. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2024.100763>
33. Bakas, K., Saddik, A., Dliou, A., Hssaisoune, M., El Hachemy, S., Ait-Ichou, H. et al. (2026). Efficient AI-Powered Orange Health Assessment And Yield Estimation. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-3-032-18857-1_30
34. Chakraborty, S., Jana, S. (2023). An Integrated Framework for Smart Monitoring of Orange Orchard. *2023 7th International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2023*. <https://doi.org/10.1109/IEMENTech60402.2023.10423538>
35. Changgeng Yu, Guo, J., Pan, Q. (2026). CWG-YOLOv8: a Navel Orange Detection Model Based on Improved YOLOv8 in an Agricultural Environment. *Optical Memory and Neural Networks (Information Optics)*. <https://doi.org/10.3103/S1060992X24601325>
36. Chen, W., Lu, S., Liu, B., Li, G., Qian, T. (2020). Detecting citrus in orchard environment by using improved Yolov4. *Scientific Programming*. <https://doi.org/10.1155/2020/8859237>
37. Chen, Z., Zhang, Y., Qin, J. (2025). SALC-Net: an efficient and accurate green citrus detection model for edge devices. *Measurement Science and Technology*. <https://doi.org/10.1088/1361-6501/ae1aa2>
38. Jing, J., Zhai, M., Dou, S., Wang, L., Lou, B., Yan, J. et al. (2024). Optimizing the YOLOv7-Tiny Model with Multiple Strategies for Citrus Fruit Yield Estimation in Complex Scenarios. *Agriculture (Switzerland)*. <https://doi.org/10.3390/agriculture14020303>
39. Jing, H., Fan, R., Hong, L., Li, Z. (2025). YOLO-ESS: Improved Citrus Recognition under Natural Orchard Conditions. *Proceedings of 2025 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC 2025*. <https://doi.org/10.1109/SDPC68151.2025.11347840>
40. Kim, Y.S., Chung, Y.S., Heo, S. (2022). Image analysis as a potential tool for marker-assisted selection. *Plant Biotechnology Reports*. <https://doi.org/10.1007/s11816-021-00740-y>
41. Li, Y., Gong, Z., Zhou, Y., He, Y., Huang, R. (2023). Production Evaluation of Citrus Fruits based on the YOLOv5 compressed by Knowledge Distillation. *Proceedings of the 2023 26th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2023*. <https://doi.org/10.1109/CSCWD57460.2023.10152740>
42. Lu, J., Chen, P., Yu, C., Lan, Y., Yu, L., Yang, R. et al. (2023). Lightweight green citrus fruit detection method for practical environmental applications. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2023.108205>
43. Lyu, S., Zhao, Y., Li, R., Chen, Q., Li, Z. (2021). The accurate recognition system of citrus flowers using YOLOv4-Tiny lightweight neural network and FPGA embedded platform. *Proceedings of SPIE - The International Society for Optical Engineering*. <https://doi.org/10.1117/12.2611299>
44. Ponce-Machete, R.A., Fajardo, A. (2024). Predicting Calamansi Fruit Yield Using CNN-Based Flower Detection: A Deep Learning Approach. *Proceedings - 21st International Joint Conference on Computer Science and Software Engineering, JCSSE 2024*. <https://doi.org/10.1109/JCSSE61278.2024.10613627>
45. Xiao, X., Jiang, Y., Wang, Y. (2024). A Method of Robot Picking Citrus Based on 3D Detection. *IEEE Instrumentation and Measurement Magazine*. <https://doi.org/10.1109/MIM.2024.10505191>
46. Zhai, M., Jing, J., Dou, S., Du, J., Wang, R., Yan, J. et al. (2025). Deriving Early Citrus Fruit Yield Estimation by Combining Multiple Growing Period Data and Improved YOLOv8 Modeling. *Sensors*. <https://doi.org/10.3390/s25154718>

47. Zheng, Z., Xiong, J., Wang, X., Li, Z., Huang, Q., Chen, H. et al. (2023). An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle. *Journal of Field Robotics*. <https://doi.org/10.1002/rob.22147>
48. Zhou, C., Xu, F., Huang, A., Wang, C., Zhang, W. (2024). Utilizing NeRF-Based Rays for Spatial Perception in Fruit Counting Deduplication. *Proceedings - 2024 International Symposium on Internet of Things and Smart Cities, ISITSC 2024*. <https://doi.org/10.1109/ISITSC64373.2024.00013>
49. Feng, Y., Ma, W., Tan, Y., Yan, H., Qian, J., Tian, Z. et al. (2024). Approach of Dynamic Tracking and Counting for Obscured Citrus in Smart Orchard Based on Machine Vision. *Applied Sciences (Switzerland)*. <https://doi.org/10.3390/app14031136>
50. Gremes, M.F., Fermo, I.R., Krummenauer, R., Flores, F.C., Andrade, C.M.G., Lima, O.C.d.M. (2023). System of Counting Green Oranges Directly from Trees Using Artificial Intelligence. *AgriEngineering*. <https://doi.org/10.3390/agriengineering5040111>
51. Zheng, Z., Wu, M., Chen, L., Wang, C., Xiong, J., Wei, L. et al. (2024). A robust and efficient citrus counting approach for large-scale unstructured orchards. *Agricultural Systems*. <https://doi.org/10.1016/j.agsy.2024.103867>
52. Vijayakumar, V., Costa, L., Ampatzidis, Y. (2021). Prediction of citrus yield with AI using ground-based fruit detection and UAV imagery. *American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2021*. <https://doi.org/10.13031/aim.202100493>
53. Vijayakumar, V., Ampatzidis, Y., Costa, L. (2023). Tree-level citrus yield prediction utilizing ground and aerial machine vision and machine learning. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2022.100077>
54. Daiaeddine, M.J., Badrouss, S., El Harti, A., Bachaoui, E.M., Biniz, M., Mouncif, H. (2025). UAV imagery, advanced deep learning, and YOLOv7 object detection model in enhancing citrus yield estimation. *Foods and Raw Materials*. <https://doi.org/10.21603/2308-4057-2025-2-650>
55. Chen, J., Liu, H., Zhang, Y., Zhang, D., Ouyang, H., Chen, X. (2022). A Multiscale Lightweight and Efficient Model Based on YOLOv7: Applied to Citrus Orchard. *Plants*. <https://doi.org/10.3390/plants11233260>
56. Zeng, Y., Lin, Y., He, Y., Li, T., Li, J., Wang, B. et al. (2024). Enhanced progressive fusion method for the efficient detection of multi-scale lightweight citrus fruits. *International Journal of Agricultural and Biological Engineering*. <https://doi.org/10.25165/j.ijabe.20241706.8802>
57. Lv, Q., Sun, F., Bian, Y., Wu, H., Li, X., Li, X. et al. (2025). A Lightweight Citrus Object Detection Method in Complex Environments. *Agriculture (Switzerland)*. <https://doi.org/10.3390/agriculture15101046>
58. Xu, L., Li, H., Tang, Z., He, Y., Yu, Y., Zi, L. et al. (2026). A multi-ripeness citrus detection model based on OR-YOLOv11. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2026.102254>
59. Yu, Y., Liu, Y., Li, Y., Xu, C., Li, Y. (2024). Object Detection Algorithm for Citrus Fruits Based on Improved YOLOv5 Model. *Agriculture (Switzerland)*. <https://doi.org/10.3390/agriculture14101798>
60. Xu, W., Ji, T., Su, Q., Zheng, J., Shen, C., Ji, C. (2026). FPH-CDet: A Lightweight Citrus Detection Method for Low-Cost Embedded Device Deployment. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2026.3693846>
61. Chen, W., Lu, S., Liu, B., Chen, M., Li, G., Qian, T. (2022). CitrusYOLO: A Algorithm for Citrus Detection under Orchard Environment Based on YOLOv4. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-12687-5>
62. Cheng, Y., Gong, R., Zhang, X., Zhang, T., Zha, X., Cao, Z. et al. (2025). Enhanced YOLO and RAFT for Real-Time Detection and Depth Estimation of Multiple Fruits in Orchard Environments. *ICCIP 2024 - 2024 the 10th International Conference on Communication and Information Processing*. <https://doi.org/10.1145/3708657.3708729>
63. Deng, T., Sun, J., Zhao, Q., Huang, F. (2026). YOLO-FSEP: An Improved YOLOv8n Algorithm for Sugar Orange Detection in Orchards. *Sensors*. <https://doi.org/10.3390/s26123848>
64. Hassan, A., Mumtaz, R., Palade, V., Amin, A., Mahmood, Z., Khan, N. et al. (2025). Orange yield estimation using object tracking and 3D reconstruction. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2025.101088>
65. Isidro, E., Corbera, M., Jimenez, G., Romero, S. (2025). Detection and Quantification of Citrus Unshiu Mandarins with YOLO: A Comparative Analysis. *Proceedings of the 2025 IEEE 32nd International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2025*. <https://doi.org/10.1109/INTERCON67304.2025.11244684>
66. Li, G., Huang, X., Ai, J., Yi, Z., Xie, W. (2021). Lemon-YOLO: An efficient object detection method for lemons in the natural environment. *IET Image Processing*. <https://doi.org/10.1049/ipr2.12171>
67. Nan, Q., Huang, W., Min, H. (2024). A Fruit Detection Model in Complex Environments Based on Attention Mechanism and Feature Fusion. *Proceedings - 2024 China Automation Congress, CAC 2024*. <https://doi.org/10.1109/CAC63892.2024.10865515>
68. Peng, H., Chen, H., Zhang, X., Liu, H., Chen, K., Xiong, J. (2024). Retinanet_G2S: a multi-scale feature fusion-based network for fruit detection of punna navel oranges in complex field environments. *Precision Agriculture*. <https://doi.org/10.1007/s11119-023-10098-6>
69. Shankarpure, M., Patil, D.D. (2026). YOLO Algorithm-Based Effective Orange Detection and Localization with Improved Data Augmentation. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-981-96-7807-5_32
70. Tang, Y., Huang, W.X., Tan, Z.P., Chen, W.L., Wei, S., Zhuang, J.J. et al. (2025). Citrus fruit detection based on an improved YOLOv5 under natural orchard conditions. *International Journal of Agricultural and Biological Engineering*. <https://doi.org/10.25165/j.ijabe.20251803.8935>
71. Tawfik, M., Badawy, W. (2025). AI-Driven Framework for Automated On-Tree Orange Sizing. *2025 International Conference on Future Telecommunications and Artificial Intelligence, IC-FTAI 2025 - Proceedings*. <https://doi.org/10.1109/IC-FTAI67960.2025.11384711>
72. Wang, X., Huang, Y., Wei, S., Xu, W., Zhu, X., Mu, J. et al. (2025). ELD-YOLO: A Lightweight Framework for Detecting Occluded Mandarin Fruits in Plant Research. *Plants*. <https://doi.org/10.3390/plants14111729>
73. Xu, L., Wang, Y., Shi, X., Tang, Z., Chen, X., Wang, Y. et al. (2023). Real-time and accurate detection of citrus in complex scenes based on HPL-YOLOv4. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2022.107590>

74. Yan, L., Liming, X., Weiqi, L., Hao, L., Jie, L. (2022). RESEARCH ON RECOGNITION OF OCCLUDED ORANGE FRUIT ON TREES BASED ON YOLOv4; [YOLOv4]. *INMATEH - Agricultural Engineering*. <https://doi.org/10.35633/inmateh-67-13>
75. Ye, C., Wang, P., Zhang, J., Wang, J., Wang, T., Liu, Y. (2023). Design of a Deep Learning-Based General Fruit Recognition Algorithm. *IEEE 1st International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics, AIKIE 2023*. <https://doi.org/10.1109/AIKIE60097.2023.10389926>
76. Zheng, Z., Xiong, J., Lin, H., Han, Y., Sun, B., Xie, Z. et al. (2021). A Method of Green Citrus Detection in Natural Environments Using a Deep Convolutional Neural Network. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2021.705737>
77. Cong, G., Chen, X., Bing, Z., Liu, W., Chen, X., Wu, Q. et al. (2025). YOLOv8-Scm: an improved model for citrus fruit sunburn identification and classification in complex natural scenes. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2025.1591989>
78. Chen, S., Xiong, J., Jiao, J., Xie, Z., Huo, Z., Hu, W. (2022). Citrus fruits maturity detection in natural environments based on convolutional neural networks and visual saliency map. *Precision Agriculture*. <https://doi.org/10.1007/s11119-022-09895-2>
79. Wang, C., Han, Q., Li, C., Zou, T., Zou, X. (2024). Fusion of fruit image processing and deep learning: a study on identification of citrus ripeness based on R-LBP algorithm and YOLO-CIT model. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2024.1397816>
80. Qu, H., Du, H., Tang, X., Zhai, S. (2025). Citrus fruit diameter estimation in the field using monocular camera. *Biosystems Engineering*. <https://doi.org/10.1016/j.biosystemseng.2025.02.012>
81. Sukkuea, A., Kurdthongmee, W., Suwannarat, K. (2026). AGRO-YOLO-V: Hybrid Instance Segmentation and Geometric Modeling for Single-View Orange Volume Estimation. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2026.3667161>
82. Calcagno, S., Midolo, A., Scaletta, E., Tramontana, E., Verga, G. (2026). Detecting Objects in Aerial Imagery Using Drones and a YOLO-C3 Hybrid Approach. *Future Internet*. <https://doi.org/10.3390/fi18040204>
83. Ilić, M., Kolarski, M., Tomčić, A., Penjivrag, S., Tošić, M. (2026). Edge-Driven Orchard Analytics: Automated Treetop Detection using YOLOv11. *2026 49th MIPRO ICT and Electronics Convention, MIPRO 2026 - Proceedings*. <https://doi.org/10.1109/MIPRO70003.2026.11591804>
84. Kahveci, S., Çelik, M.Ö., Akkurt, R., Kahveci, Ö. (2025). Deep learning based citrus tree detection from low resolution satellite images: A case study of Tarsus. *Turkish Journal of Remote Sensing*. <https://doi.org/10.51489/tuzal.1669616>
85. Zhang, L., Bao, Z., Chen, H., Kang, H., Li, W., Chen, Y. (2026). InspectGaussian: Large-scale coarse-to-fine Gaussian reconstruction for orchard inspection robots. *Plant Phenomics*. <https://doi.org/10.1016/j.plaphe.2026.100195>
86. Zhang, L., Luo, P., Ding, S., Li, T., Qin, K., Mu, J. (2024). The grading detection model for fingered citron slices (citrus medica 'fingered') based on YOLOv8-FCS. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2024.1411178>
87. Li, Z., Zhou, Y., Zhao, C., Guo, Y., Lyu, S., Chen, J. et al. (2023). Design of a Cargo-Carrying Analysis System for Mountain Orchard Transporters Based on RGB-D Data. *Applied Sciences (Switzerland)*. <https://doi.org/10.3390/app13106059>
88. Peng, H., Li, W., Xie, H., Liu, H., Li, X. (2026). Edge-YOLOv11: Lightweight UAV fruit detection in dense canopies for precision agriculture. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2026.102133>
89. Wang, S., Weng, L., Hao, L., Zuo, S., Yang, Z., Wang, M. et al. (2026). Design and application of multimodal visual-tactile sensor for object information perception. *Sensors and Actuators A: Physical*. <https://doi.org/10.1016/j.sna.2026.117529>
90. Zhou, C., Ampatzidis, Y., Guan, H., Liu, S., Liu, W., de Oliveira Costa Neto, A. et al. (2025). Agrosense: Accelerating precision orchard management through an AI-enabled monitoring system. *Precision Agriculture*. <https://doi.org/10.1007/s11119-025-10268-8>
91. Carella, A., Lucas, B.P.E., El Ghazouali, S., Bulacio Fischer, P.T., Massenti, R., Venturini, F. et al. (2026). Large-scale orange fruit dataset for localization, classification and ripening assessment under varying environments. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2026.111833>
92. Saritas, M.M., Golcuk, A., Koklu, M. (2026). Detection of Oranges on Trees Across Diverse Field Conditions Using YOLOV11 Models. *Applied Fruit Science*. <https://doi.org/10.1007/s10341-026-01810-7>
93. Montalban-Faet, G., Navarro-Modesto, E., Salcedo-Navarro, A., Fayos-Jordan, R., Benlloch-Caballero, P., Garcia-Pineda, M. et al. (2026). Dataset for orange fruit detection from UAV in citrus orchards. *Data in Brief*. <https://doi.org/10.1016/j.dib.2026.112733>
94. Khedkar, M., Sambhe, V., Dhore, V., Khedkar, S., Khandare, N. (2025). OrangeFruitNet: A Dataset and On-device Real Time YOLO-based Framework for Orange Fruit Detection and Yield Estimation in Orchard Environments. *International Journal of Intelligent Engineering and Systems*. <https://doi.org/10.22266/ijies2025.1231.67>
95. Beldek, C., Sariyildiz, E., Alici, G. (2025). A Deployment-Oriented Benchmarking of You Look Only Once (YOLO) Models for Orange Detection and Segmentation in Agricultural Robotics. *Agriculture (Switzerland)*. <https://doi.org/10.3390/agriculture15202170>
96. Jroni, Z., Moussaid, A., Hadi, M.Y. (2024). Exploring End-to-End object detection with transformers versus YOLOv8 for enhanced citrus fruit detection within trees. *Systems and Soft Computing*. <https://doi.org/10.1016/j.sasc.2024.200103>
97. Zheng, S., Lin, H., Liang, J., Yang, M. (2024). Research on Citrus Object Detection Method Based on Improved YOLOv5. *2024 4th International Symposium on Computer Technology and Information Science, ISCTIS 2024*. <https://doi.org/10.1109/ISCTIS63324.2024.10699204>
98. Toscano, F., D'Antonio, P., Santana, L.S., Fiorentino, C. (2026). Olive Yield Prediction in the Mediterranean Basin: Bibliometric Evidence of Precision Agricultural Engineering Gaps and Innovation Priorities for Sustainable Agri-Food Systems. *Agronomy*, 16, 1189. <https://doi.org/10.3390/agronomy16121189>

Figures

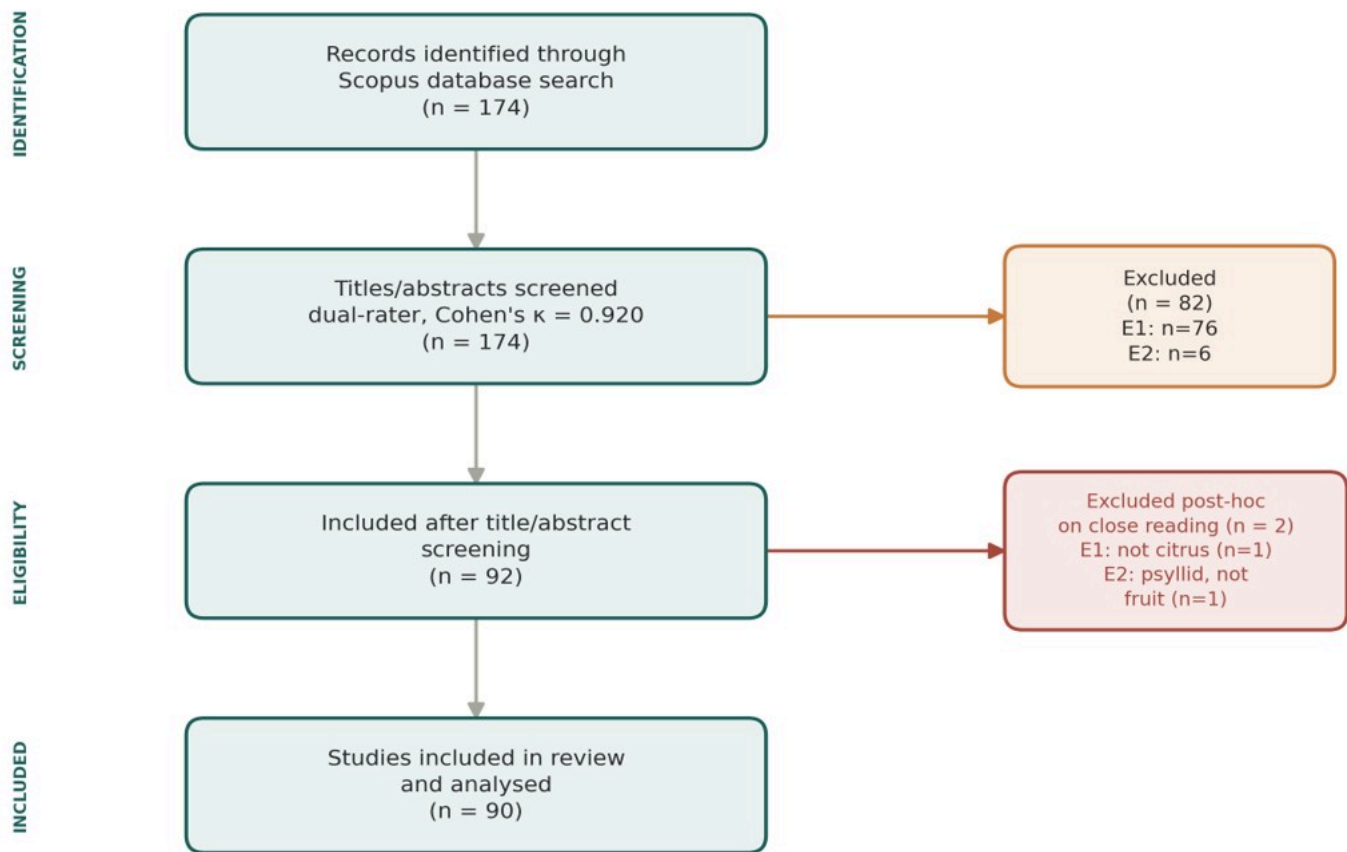


Figure 1

PRISMA 2020 flow diagram showing the post-hoc exclusion (§2.2) as a distinct step from title/abstract screening

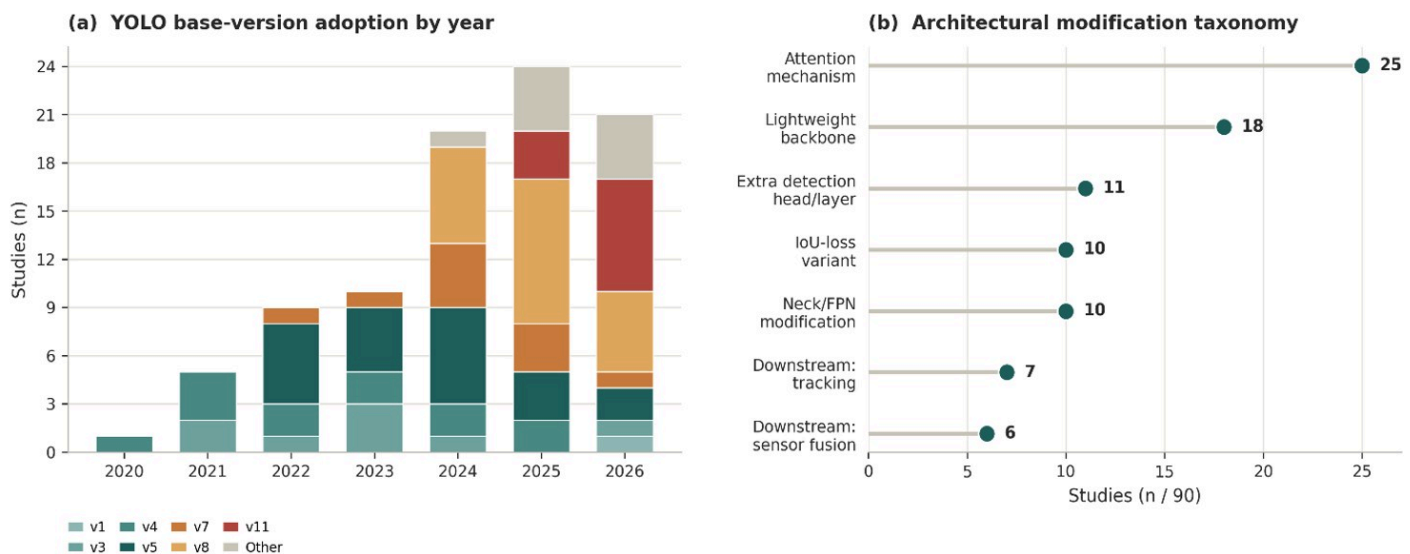


Figure 2

a YOLO base-version adoption by publication year; **b** frequency of architectural/pipeline modification categories (studies may use more than one category)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytableS1.csv](#)
- [riskofbiasassessment.csv](#)
- [supplementarytableS2.csv](#)