Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 5 (2026) 588

African Scientific Reports

Smart automation through computer vision and image processing: a systematic literature review

Osita Miracle Nwakeze[a,*], Nwabudike Uju Cynthia[b], Omorogie Michael[b], Chukwura Anthony Afamefuna[b], Udensi Nkechi Grace Cecilia[c], Ezekiel-Odimgbe Chinenye Love[d]

[a]*Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria*
[b]*Department of Computer Science, Delta State Polytechnic, Ogwashi-Uke, Delta State, Nigeria*
[c]*Department of Computer Science Education, Nwafor Orizu College of Education, Nsugbe, Anambra State, Nigeria*
[d]*Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*

Abstract

This systematic literature review focused on the role of computer vision (CV) and image processing (IP) in smart automation, including their applications in manufacturing processes, healthcare, agricultural production, infrastructure, and edge computing, as well as their implications for these areas. The OpenAlex database was used to identify relevant studies published between 2013 and 2025. Eight empirical studies were included in this review, each reporting quantitative results. The reviewed articles evaluated the accuracy of classification, average precision, throughput, and various rates of defects detected for specific purposes, such as inspection, pest detection, parking, edge vision, and simulated instrument use. Due to the differences in their findings and contexts, it was not possible to combine the results reported by the selected studies into a single summary estimate. The reviewed articles indicate that computer vision techniques can be utilized to facilitate certain forms of automated decision-making but require additional research into external validity and actual performance after implementation. This article summarizes the eight studies included in the completed OpenAlex search.

DOI: 10.46481/asr.2026.5.3.588

Keywords: Computer vision, Image processing, Smart automation, Deep learning, Artificial intelligence

Article History:
Received: 18 June 2026
Received in revised form: 06 August 2026
Accepted for publication: 20 October 2026
Available online: 04 November 2026

© 2026 The Author(s). Published by the Nigerian Society of Physical Sciences under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

The rise of Industry 4.0, artificial intelligence (AI), and the Internet of Things (IoT) has enabled the connection of automated equipment with data-driven monitoring and control [1]. Conventional automation systems often make use of fixed rules and carefully specified input conditions. Alterations in lighting, object position and operating settings may involve the need for visual feedback and adjustment. Computer vision and image processing enable automated machines to observe the environment, identify relevant characteristics and supply information for a decision in an operation. The practical value of these approaches depends on the accuracy of perception and how reliably the full system makes use of their output in real operating circumstances.

*Corresponding Author Tel. No.: +234-802-207-0309.
Email address: ma.nwakeze@coou.edu.ng (Osita Miracle Nwakeze)

Image processing enhances or evaluates the visual data captured by an automation system through operations such as filtering, segmentation and feature extraction. Computer vision uses those data to identify objects, recognise patterns and track events. Both approaches support tasks such as industrial inspection, medical image analysis, crop monitoring and traffic management. Performance may vary depending on lighting, camera position or the objects being observed. Deep learning can learn visual representations from examples, but its practical value still depends on representative data and tests conducted under relevant operating conditions [2, 3].

Convolutional neural networks (CNNs) and related deep learning models can learn image features directly from training data [3]. This reduces the need to specify every visual feature by hand, although preprocessing, data quality, and validation remain important. Combining image processing, computer vision and a clearly defined decision process can increase the flexibility of an automation system. This paper examines these methods in industrial, healthcare, agricultural and urban applications, compares the evidence available for their performance and identifies questions that remain for implementation.

2. Literature review

Computer vision and image processing support automation by enabling machines to interpret and respond to visual stimuli. Cheng *et al.* [4] investigated graphical user interface (GUI) automation in industrial applications, and described their Korat system, which used image-based interaction and optical character recognition across disparate interfaces. Cheng [5] and Cui and Li [6] discussed computer vision in industrial work, including workpiece orientation recognition. Park and Jeong [7] implemented computer vision inspection for masks, while Silva and Paladini [8] applied it to an automotive assembly line. The tasks differ substantially: performance on interface screenshots cannot substitute evidence of defect detection in varying light and production conditions.

More and More [9] described a conveyor sorting system incorporating computer vision and IoT communication, illustrating a production workflow that reacts to camera observations. Rahman *et al.* [1] reviewed machine learning in the context of industrial automation, including the roles of edge AI and digital twins, while Yousif *et al.* [10] examined computer vision in the context of autonomous industrial facilities. These works provide context for linking visual inference to equipment and process control [4], and should not be regarded as independent empirical studies within the scope of this eight-study dataset. In these included studies, Chou and Liu [11] applied visual recognition to an operational truck-entry audit, and Park and Jeong [7] reported an increase in mask production throughput following the introduction of a broader inspection system. As equipment and process changes were also applied, the latter before-and-after result does not isolate the contribution of the vision model.

The integration of robotics and computer vision has also been investigated in the context of smart manufacturing [12]. Bhana *et al.* [13] presented a vision-based approach to industrial safety, with attention to personal protective equipment and workplace activity, while this conference study provided contextual evidence but was not among the eight papers identified through this specified indexed search. Mira *et al.* [14] compared conventional and convolutional classifiers for olive-fly detection in terms of accuracy and edge energy use, while Li *et al.* [15] evaluated a modified YOLO detector in pheromone-baited pest monitoring. Such applications demonstrate the importance of deployment conditions and the reported outcome, alongside the choice of model family. Zhou *et al.* [16] reviewed generative AI in the context of industrial computer vision, encompassing augmentation and anomaly detection; as a secondary review, it contributes background rather than an additional independent result. These sources motivate a comparison distinguishing empirical outcomes, the domain conditions within which operations take place, and the operational action and remaining evidence gaps.

3. Research methodology

This systematic literature review followed the relevant Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [17]. A search conducted on 27 September 2026 utilized three OpenAlex /works queries, covering articles published between 1 January 2013 and 31 December 2025; the common filter was from 2013 to 2025. The query fields were S1, title.search: computer vision and abstract.search: smart automation; S2, title.search: machine vision and abstract.search: smart automation; and S3, title.search: image processing and abstract.search: smart automation. The queries returned 33, 20 and 26 records, respectively; records were exported without a relevance cutoff, and the search audit includes the queries, metadata and selection decisions.

3.1. Eligibility, screening and appraisal

The reviewed articles were selected based on inclusion criteria requiring a visual input used in a smart automation decision or action and an assessable quantitative outcome. Thus, secondary reviews, theses and studies that lacked a relevant visual task or assessable outcome were excluded. Duplicates were identified using Digital Object Identifiers (DOIs). Reports that did not have accessible full text via the documented retrieval route were recorded as not retrieved. The selection process and its counts are presented in Figure 1.

The inclusion and exclusion criteria are summarized in Table 1.

Table 1: Inclusion and exclusion criteria.

Dimension	Included	Excluded or not retrieved
Record	Article dated 2013–2025 with an empirical visual automation application	Secondary work or report outside the date range
Application	Visual input supports an automated decision or action	No relevant visual automation task
Outcome	Assessable quantitative outcome evidenced at screening and checked in the available report	No outcome evidenced in the indexed abstract
Retrieval	Full text available through the documented indexed PDF route	Full text unavailable through the documented route

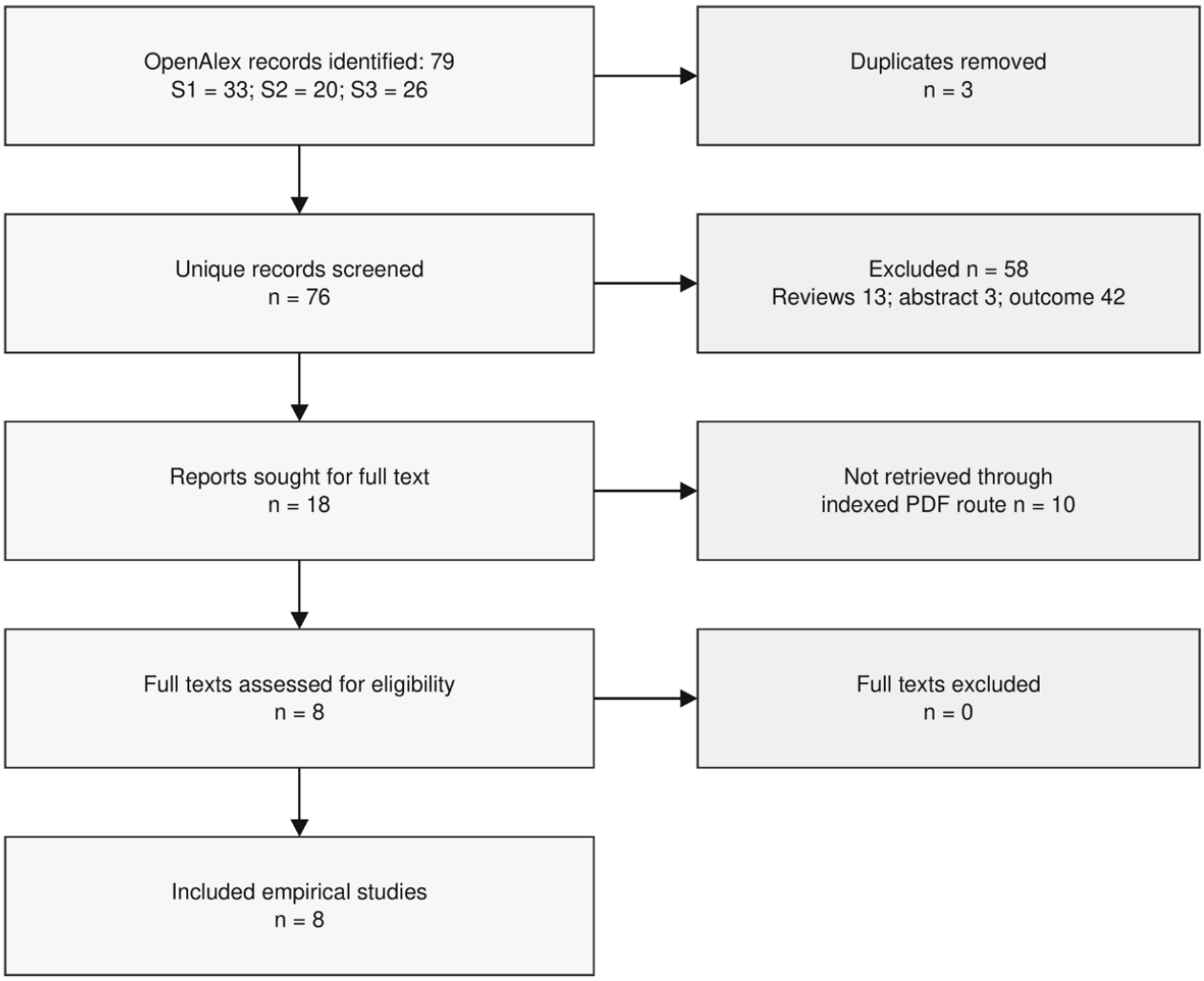

Figure 1: PRISMA flow diagram for the completed OpenAlex review.

3.2. Search results and study selection

The OpenAlex queries yielded 79 records (33 + 20 + 26). After three DOI duplicates were removed, 76 unique records were screened. Fifty-eight were excluded during title and abstract screening: 13 secondary works, three without assessable indexed abstracts, and 42 without an outcome evidenced in the abstract. Eighteen full texts were sought; ten were not retrieved through the documented route. The remaining eight reports were assessed and included. Figure 1 shows these decisions.

Table 2: Characteristics and evidence extracted from the eight included studies.

Study	Application	Dataset	Model	Reported outcome	Limitation
Chou and Liu [11]	Truck-plate audit	Site plate images	YOLOv3 + convolutional character recognition	93.73% overall recognition; 0.3271 s/image	Single operating setting
Mira *et al.* [14]	Olive-fly monitoring	3,043 regions from four farms	PCA + random forest; HOG + SVM; LeNet-5 CNN	Best 99% classification accuracy	Class imbalance and field variation
Park and Jeong [7]	Mask inspection	Production-line mask images	Vision system with proprietary training toolkit	2,600→4,189 sheets/hour; defect rate 8.3%→5.69%	Before–after system change
Silva and Paladini [8]	Assembly inspection	Line-camera components and screw threads	Hybrid machine vision sensor	Production-run inspection; no comparable classifier accuracy	Specific production setting
Li *et al.* [15]	Pest monitoring	Seven target species in trap images	YOLOv9-TrapPest	97.5% precision; 98.3% mAP50	Limited deployment diversity
Shaushenova *et al.* [18]	Edge person detection	Visual Wake Words benchmark	Hybrid CNN–KAN	82.32% classification accuracy	Benchmark rather than field action
Rani *et al.* [19]	Parking automation	3,500 images; 70/30 split	YOLOv5 + image processing	97.11% without spaces; 91.91% with spaces	Camera and condition dependence
Shreyas *et al.* [20]	Surgical tool handling	4,261 instrument images; simulated setting	VGG16 + voice control	>95% tool ID; 92.5% voice recognition	No clinical outcome validation

3.3. Data extraction and quality assessment

For the eight included studies, the available full texts were examined for application, visual data, model, reported outcome and study-specific concerns. Table 2 presents the extracted characteristics. Accuracy, mean average precision, throughput and defect rate were treated as distinct outcomes. Table 3 describes study design and reporting concerns.

4. Computer vision, image processing and applications

Computer vision (CV) and image processing (IP) are two related techniques that enable machines to acquire, improve, interpret and utilize visual information. Often, the image is modified or measured (e.g., by filtering, thresholding, or segmenting). Computer vision is the process of extracting information about the objects, regions, motion or events in the image. In practice, these functions overlap and either or both can play a role in making an automated decision [2]. The automation chain also includes sensing, software integration, a control or human review step and feedback from the action.

Image processing is used to prepare the image for analysis by reducing noise, correcting contrast, finding regions of interest, and extracting image features. When the scene is constrained, classical edge detection, morphological operations and histogram transformations might be effective [2]. The method used will be determined by the camera location, lighting, material and error costs. It is also possible for a processing step to lose valuable information; the impact of the processing step should be evaluated with reference to the final task, not the appearance of the product. Under the edge conditions tested, Mira *et al.* [14] found that feature-based classifiers and a convolutional network were effective when applied to images of pests.

Computer vision is the process of assigning meaning to the visual data by using object recognition, detection, tracking, segmentation and scene analysis. Deep convolutional neural networks learn the visual features from examples, and residual connections have been exploited for this purpose [3]. The family of vision transformers generalizes this model class with attention on image patches [21]. These architectures can be effective but their performance relies on the quality and representativeness of training images as

Table 3: Study design and reporting concerns in the eight included studies.

Study	Study design / evidence type	Reporting strengths	Principal validity concern
Chou and Liu [11]	Site-specific plate-recognition test	Dataset and recognition/time outcomes described	Transfer to other camera sites and conditions untested
Mira *et al.* [14]	Four-farm trap-image comparison on edge hardware	Model comparators, runtime and energy reported	Site/season transfer and class balance need separate checks
Park and Jeong [7]	Before–after mask-line implementation	Throughput and defect outcomes reported	Equipment and workflow changes confound algorithm effect
Silva and Paladini [8]	Assembly-line system case study	Production setting and inspection workflow described	No isolated model baseline or comparable error estimate
Li *et al.* [15]	Trap-image detection experiment	Train/validation/test split and detector comparisons reported	Generalization and on-trap runtime not established
Shaushenova *et al.* [18]	Visual Wake Words benchmark	Named dataset and computational experiments	Benchmark result does not measure field actuation
Rani *et al.* [19]	Parking-image test	Two reported operating conditions	Camera/site dependence and external validation unclear
Shreyas *et al.* [20]	Simulated surgical-tool prototype	Tool and voice recognition reported	No clinical outcome or independent patient-setting test

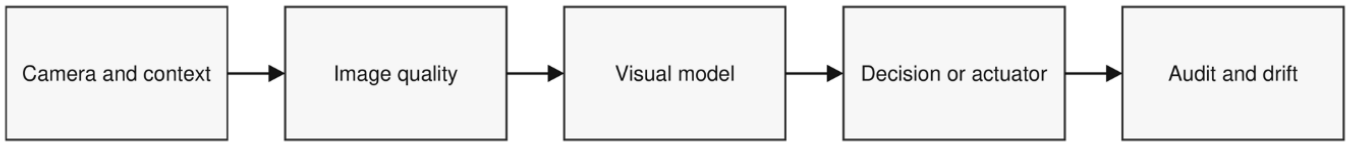

Figure 2: Conceptual framework.

well as the separation of training and test data and operating environment. Model performance should thus be reported along with the dataset, task, metric, and the conditions under which the inferences were made.

Applications of computer vision and image processing in manufacturing inspection [7, 8], agricultural pest monitoring [14, 15], vehicle auditing and parking [11, 19] and handling of healthcare instruments [20] have been studied. These examples measure a variety of different outcomes. The recognition rate for vehicle plates, product defect rate, mean average precision for pest detection and simulated surgical-tool identification accuracy are answers to different questions. Furthermore, a high image-level result is not necessarily a sign of safe and reliable operation if the system operates autonomously. Although edge computing can reduce camera-to-decision latency, the latency and energy reduction achieved will be dependent on the specified hardware [14, 18].

Visual sensing and automatic action consequently provide opportunities in various fields and require careful validation. A real system needs to specify what the camera sees, what the model tries to predict, when the operator steps in and what errors are observed after the system is deployed. The relationship between image-processing, visual-inference, action, and feedback stages is shown in Figure 2. The main application domains of the reviewed studies are grouped in Figure 3.

4.1. Industrial automation

Computer vision and image processing can be used in industrial quality control to detect deviations from the expected appearance of products, missing parts and defects. Conventional image processing, convolutional classifiers and real-time object detectors are examples of methods. Park and Jeong [7] reported that the installed mask-inspection system achieved an increase in production

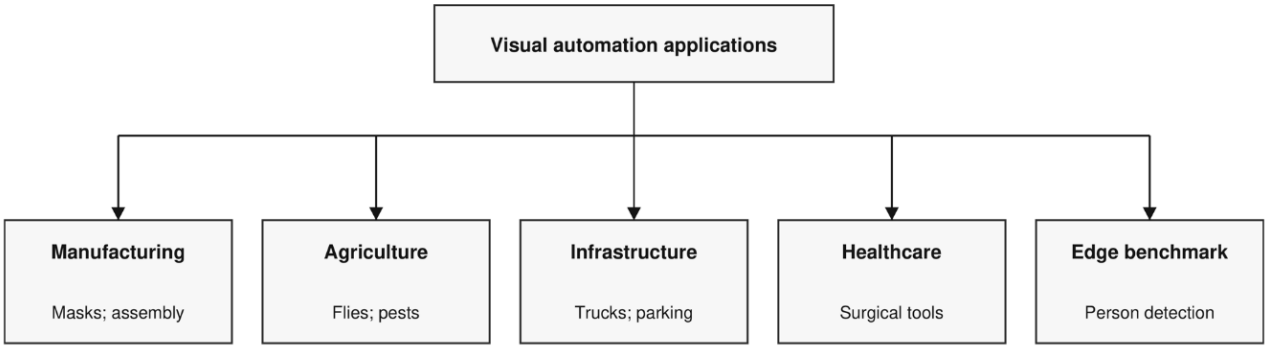

Figure 3: Application taxonomy.

rate from 2,600 sheets/hour to 4,189 sheets/hour, and that the reported defect rate dropped from 8.3% to 5.69%, a before-and-after system result rather than an isolated model-accuracy measurement. Silva and Paladini [8] introduced a vision-based inspection system for components and screw threads in assembly. Some earlier industrial research into GUI automation [4] and assembly visual recognition [6] addresses related but different tasks. Digital-twin integration can link the visual measurements to the process representation [1, 10] but synchronisation and field validation should be documented.

4.2. Healthcare automation

In healthcare, image processing and computer vision can be used to aid imaging, patient monitoring and robotic workflows. Radiographic, computed tomography and magnetic resonance analyses should be evaluated using the relevant patient population and independent clinical data to assess the automated analysis of these images. In the included study, Shreyas *et al.* [20] developed an image classifier based on the VGG16 model and a voice-controlled robot to recognize and manipulate surgical instruments in a simulated environment, achieving over 95% accuracy in recognizing the tools and over 92.5% accuracy in voice recognition. These measures do not provide clinical safety or patient outcomes. Human supervision, failure modes, and calibration and workflow impact are therefore important considerations for vision-guided systems, and should be evaluated before any autonomous clinical use. Pak *et al.* [22] carried out a separate, publicly available comparative study of the image segmentation during robot-assisted prostatectomy. Three convolutional models, DeepLabv3, MANet, and U-Net++ had the average Dice scores of 0.938, 0.944 and 0.930, respectively, whereas three transformer models, SegFormer, BEiT, and DPT had the average Dice scores of 0.919, 0.916 and 0.940, respectively. This study compares three convolutional models and three transformer models on the same surgical image segmentation task. This study is not included in the eight-study evidence base and is included only as contextual literature.

4.3. Agriculture and environmental automation

Cameras mounted on traps, equipment or other platforms can be used to monitor crops, pests and livestock, which can help to reduce some manual monitoring in precision agriculture. Mira *et al.* [14] tested HOG support vector machine, LeNet-5 convolutional network and PCA random forest classifiers for the recognition of olive flies. They extracted data for 3,043 image regions from four farms, the best reported classification accuracy was 99% with favourable edge-processing time and energy use. Li *et al.* [15] created a monitoring system using pheromone bait and developed a model based on the YOLOv9 algorithm for seven pest species and achieved a mean average precision of 98.3% at an intersection-over-union (IoU) threshold of 0.5. Validation across seasons and locations are still relevant; these results rely on specific designs of traps, images and test conditions. Automated irrigation or fertilisation is a separate task of control, and cannot be inferred from pest classification alone.

4.4. Smart cities and safety systems

Visual data is used in urban systems and public safety applications for vehicle recognition, parking operations and workplace monitoring. Chou and Liu [11] examined an automatic truck-plate recognition system for river-dredging operations with an average overall recognition rate of 93.73% and an average processing time of 0.3271 seconds per image. Rani *et al.* [19] combined YOLOv5 and image processing in an intelligent parking system. They reported an accuracy rate of 97.11% without consideration of parking spaces and 91.91% when parking spaces were considered; these measures should not be conflated. Bhana *et al.* [13] concerned with personal protective equipment for the industry, and did not assess urban traffic quality. Issues of occlusion, optical degradation, privacy compliance, fairness and error escalation need to be evaluated before deployment in unconstrained urban environments.

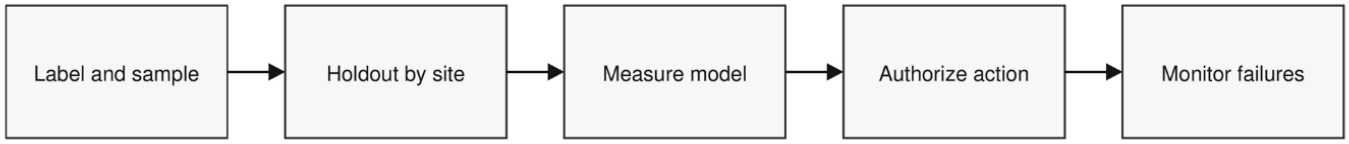

Figure 4: Deployment evaluation workflow.

Table 4: Descriptive publication and application distribution of included studies.

Year	Studies	Application	Studies
2021	1	Manufacturing	2
2022	1	Agriculture	2
2023	0	Infrastructure and parking	2
2024	3	Healthcare workflow	1
2025	3	General edge vision	1

Table 5: Primary algorithm families in the eight included studies.

Model family	Count
YOLO variants	3
PCA + random forest	1
Proprietary vision toolkit	1
Hybrid vision sensor	1
Hybrid CNN–KAN	1
VGG16	1

The reviewed research papers confirm the feasibility of visual sensing in certain automated processes, but do not establish consistent visual-sensing performance across industry, health, agriculture and cities. They use different datasets, metrics, equipment, and operating environments. Any synthesis must, therefore, include a description of what was measured and under what conditions, and acknowledge untested generalisation and safety requirements.

4.5. Comparative and quantitative synthesis

Eight empirical studies met the review criteria and formed the evidence base for synthesis. They represent two manufacturing applications, two agricultural applications, two infrastructure or parking applications, one simulated healthcare workflow and one general edge vision benchmark. Publication counts were one each in 2021 and 2022, none in 2023, and three each in 2024 and 2025. Three studies used YOLO variants; the others used principal component analysis with random forest, a proprietary inspection toolkit, a hybrid vision sensor, a convolutional neural network with a Kolmogorov–Arnold network, or VGG16. Seven used study-specific image collections; one used Visual Wake Words. These counts describe the included studies only.

The publication and application distribution is shown in Table 4 , and the primary algorithm families are shown in Table 5.

Table 6 compares primary outcomes with data provenance, processing measurements, hardware and the reported operational action within four task groups. For deployment interpretation, the relevant fields are processor and accelerator type, batch size and quantization, end-to-end camera-to-decision latency, energy per image and downstream actuation delay. Mira *et al.* [14] reported 9.05 ms and 0.0256 J per classified region for the principal-components and random-forest configuration on a Raspberry Pi 4; these are model processing measurements, not an end-to-end actuation time. The other studies do not report the same complete set of deployment measurements, so missing values are labelled individually rather than inferred from training hardware. “Not reported” means that a directly comparable value was not available in the extracted study; a server-side training accelerator is not treated as evidence of edge deployment. Mean average precision at an intersection-over-union threshold of 0.50 (mAP50) is a detection measure and cannot be compared numerically with classification accuracy, throughput or defect rate. Pooling is inappropriate without common tasks, denominators, uncertainty estimates and evaluation settings. In the mask-inspection before-and-after comparison [7],

equipment and workflow changes may contribute to throughput and defect-rate differences; the study does not isolate the causal effect of the vision algorithm.

5. Open research directions

Eight empirical studies were found that met the criteria for review and that served as the evidence for synthesis. Their results identify questions requiring further investigation before wider deployment. The following directions distinguish findings from the included studies from proposed research; other cited work provides context only.

5.1. Generalization and model adaptability

The studies included were conducted on specific data and conditions: Mira *et al.* [14] on trap images from four farms, Li *et al.* [15] on a specific pest-trap system and Shreyas *et al.* [20] on a simulated surgical workflow. These studies are not designed to set a performance standard for various sites, camera angles, seasons, or clinical settings. A useful next step is to test a fixed model on independently collected data from new locations and time periods, and report errors by class and any required adaptations. The potential benefits of domain adaptation and self-supervised learning remain to be tested beyond these eight studies.

5.2. Real-time efficiency and edge-level deployment

Most of the studies included measured processing speed and energy in different ways. Chou and Liu [11] reported that their system had a processing time of 0.3271 seconds per image for truck-recognition, and Mira *et al.* [14] compared the accuracy of the classification with the processing time and power consumption on edge hardware. Shaushenova *et al.* [18] analyzed a resource-constrained vision benchmark but did not consider an actual production line. Task accuracy should be included in future evaluations, as should be the device type, image resolution, batch size, end-to-end latency, and power consumption. If the same hardware were used, it would be possible to determine whether a smaller model achieved the response time required for each application.

5.3. Dataset availability and annotation limitations

Seven of the eight studies used study-specific image collections; the named Visual Wake Words benchmark was used by Shaushenova *et al.* [18]. Mira *et al.* [14] reported 3,043 image regions from four farms and Shreyas *et al.* [20] provided 4,261 images of surgical tools in a simulated environment. This evidence does not provide an indication of the number of open datasets across the broader area of the search set. Future research should report the sources of images and the methods used to create annotations, share appropriate datasets where permissions allow, and evaluate if models developed on one site can be deployed on another site. Potential methods include synthetic images, semi-supervised learning, and federated learning, with independent real-world data retained for testing.

5.4. Robustness and security in automation systems

The studies included here identify several variations in visual input that can occur during operation. Mira *et al.* [14] evaluated trap images captured under different deployment settings, and Rani *et al.* [19] noted limitations in deployment of the camera as a result of poor image quality. Chou and Liu [11] conducted an analysis in a specific truck-auditing setting. None of these results establishes resistance to adversarial perturbations or cyberattacks. Future research should investigate the error rates for specified lighting changes, occlusion, camera motion and image corruption, and uncertainty-based escalation to a human operator. Security testing must be conducted explicitly rather than inferred from classifier accuracy.

5.5. Integration of generative and hybrid AI models

Zhou *et al.* [16] surveyed the generative approaches for computer vision in industry, such as data augmentation and anomaly detection. The eight empirical studies included in the present search did not directly test a diffusion or generative adversarial model against a non-generative baseline in an actual automation workflow. A useful experiment would keep the real external test set constant and compare rare-error performance, inference cost, and calibration between systems trained with and without generated examples. Before it can be said that generative augmentation helps with generalisation or diminishes bias, such a comparison must be made.

5.6. Human–robot collaboration and ethical considerations

The surgical-tool prototype in [20] was simulated, and the assembly and mask-inspection studies [7, 8] presented application-specific implementations. These reports do not measure the impact of operator intervention on safety, workload and error recovery in different settings. Collaborative robotics is discussed as contextual literature by Addula and Tyagi [12]. Future field evaluations should define when an operator reviews uncertain outputs, record actions taken in response, and incorporate operator-centred measures. An analysis of privacy and fairness should be conducted when identifiable people are being captured, particularly in healthcare settings and in shared public spaces.

Table 6: Task-level outcomes and deployment characteristics of the eight included empirical studies.

Study / workflow	Data and model	Primary result	Latency / platform	Operational action and comparison limit
Chou and Liu [11] / truck audit	5,419 site images (542 test); YOLOv3 and character recognition	93.73% overall plate recognition	0.3271 s/image; Intel i7 and RTX 2080 Ti	Truck plate audit; one site and on-premises accelerator
Rani *et al.* [19] / parking	3,500 images; YOLOv5 plus image processing	97.11% without spaces; 91.91% with spaces	Not reported; platform not established	Parking status; camera conditions affect transfer
Mira *et al.* [14] / olive fly	3,043 regions, four farms; PCA + random forest, HOG + SVM, LeNet-5	Up to 99% classification accuracy	PCA + RF: 9.05 ms/region and 0.0256 J; Raspberry Pi 4	Trap-based monitoring; edge measurements depend on model
Li *et al.* [15] / pest trap	1,796 trap images, seven pest types; YOLOv9-TrapPest	98.3% mAP50; 97.5% precision	Inference time not reported; RTX 1080 Ti for development	Automatic trapping, killing and counting; model latency on trap not quantified
Park and Jeong [7] / masks	Production-line mask images; machine-vision inspection toolkit	2,600 to 4,189 sheets/h; defects 8.3% to 5.69%	Per-image latency not reported; edge devices specified	Quality inspection; before/after equipment and workflow confounding
Silva and Paladini [8] / assembly	Line-camera screw/thread images; hybrid vision sensor	Production-run inspection; no common accuracy estimate	Per-image latency not reported; assembly fixture	Inspection decision; specific line and components
Shaushenova *et al.* [18] / edge benchmark	Visual Wake Words; hybrid convolutional–Kolmogorov–Arnold network	82.32% classification accuracy	Timing evaluated by batch/resolution; NVIDIA T4 training	Person-detection benchmark; no physical actuation tested
Shreyas *et al.* [20] / surgical tools	4,261 instrument images; VGG16 and voice interface	>95% tool identification; 92.5% voice recognition	Comparable latency not reported; simulated setup	Tool request/handling simulation; no patient outcome

5.7. *Digital twin integration and cyber-physical synchronization*

Yousif *et al.* [10] discuss computer vision in autonomous industrial facilities, and Rahman *et al.* [1] discuss integration with digital twins and edge artificial intelligence. None of the eight included papers assessed an entire visual-to-digital-twin control loop or measured synchronisation error. This is a gap within the reviewed evidence base. Future research might include a comparison of the timing and accuracy of the camera observations, twin-state update and control actions in a real facility, along with the behavior of the system in the event of time delay or missing data.

5.8. *Sustainable and scalable automation systems*

Mira *et al.* [14] recorded energy use, processing time, and accuracy for olive-fly classification. Shaushenova *et al.* [18] reported model accuracy and size for a resource-constrained person-detection benchmark. The other studies were not comparable in terms of energy measurements and so the review cannot estimate the environmental cost of visual automation across domains. Future evaluations should report power, processing time, model accuracy, and service life for specified devices. These measurements would enable a fair comparison of more elaborate and simpler methods for a particular application.

In the eight reports included, there is a shared methodological requirement to evaluate results beyond the study's own images and report operationally relevant outcomes. Suggested priorities are external-site tests [14, 15, 20], measured latency and energy

[11, 14, 18], camera-condition robustness [14, 19], and clearly defined human oversight [7, 8, 20]. In the bounded evidence set, research questions remain regarding explainability, privacy-preserving learning, digital twins and generative models. The claims about effectiveness in those areas are more dependent on new comparative results than on the results of the studies summarized in this paper.

5.9. Future horizons and architectural roadmap

Vision Transformers (ViTs) [21], Contrastive Language–Image Pretraining (CLIP) [23], the Segment Anything Model (SAM) [24], and self-supervised Distillation with No Labels (DINO) [25] are examples of foundation-model approaches that could be evaluated for smart automation. The eight included studies did not directly compare these approaches in a physical visual sensing and automated action workflow. Comparative studies should measure task performance alongside latency, energy use and the consequences of incorrect decisions under specified deployment conditions.

6. Limitations

A systematic literature review (SLR) was conducted using one database (OpenAlex) and eight studies were included. One reviewer screened and extracted the evidence without independent verification, and no formal risk-of-bias scores were assigned. The studies were too diverse to be suitable for meta-analysis. The screening process for the studies included in this review changed over time in the process of refining the review protocol, so the final version of the screening process is reflected in the studies included. No supplementary records from PubMed or IEEE Xplore were used.

7. Conclusion

This systematic literature review aimed to consolidate the findings of eight empirical studies on the application of computer vision and image processing in the field of smart automation in manufacturing, healthcare, agriculture, infrastructure and edge computing. Eight empirical studies met the eligibility criteria and formed the basis for synthesis. Their results apply to specific tasks and operating conditions and cannot be aggregated into a single performance estimate.

The studies in this set used object detection, image classification, and combined image-processing techniques for mask and assembly inspection [7, 8], pest monitoring [14, 15], parking and truck recognition [11, 19], and an edge vision benchmark [18] for object recognition. Recognition accuracy, average precision, throughput, and defect rate are distinct measures and have different denominators. The studies included in this evidence-based literature review indicate that under the conditions tested, visual input can be used to inform defined automation tasks.

The results indicate the need for further testing at independent sites, systematic monitoring of operational failures, assessment of resource use, and human oversight. The practical interpretation of the observed results would be clarified with these steps.

The studies reviewed reveal applications tied to particular cameras, images, equipment, and evaluation settings. Future empirical research across different locations could assess the entire sensing-to-decision process and help substantiate broader deployment claims.

Data availability

Original search audit documents the dated OpenAlex queries, metadata, duplicate decisions and record-level selection decisions. No information was gathered at the level of the participants or images for this review. Full text of the publications is not redistributed by the publisher.

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 manuscript.

Funding

The authors received no specific funding from any public, commercial, or not-for-profit organisation for this research.

References

[1] M. A. Rahman, M. F. Shahrior, K. Iqbal & A. A. Abushaiba, "Enabling intelligent industrial automation: a review of machine learning applications with digital twin and edge AI integration", Automation **6** (2025) 37. https://doi.org/10.3390/automation6030037.

[2] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed., Springer, Cham, Switzerland, 2022. https://doi.org/10.1007/978-3-030-34372-9.

[3] K. He, X. Zhang, S. Ren & J. Sun, "Deep residual learning for image recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770. https://doi.org/10.1109/CVPR.2016.90.

[4] Y.-P. Cheng, C.-W. Li & Y.-C. Chen, "Apply computer vision in GUI automation for industrial applications", Mathematical Biosciences and Engineering **16** (2019) 7526. https://doi.org/10.3934/mbe.2019378.

[5] Q. Cheng, "Application of computer vision technology in industrial automation", Journal of Physics: Conference Series **2037** (2021) 012015. https://doi.org/10.1088/1742-6596/2037/1/012015.

[6] C. Cui & Y. Li, "Application of computer visual recognition in industrial automation assembly production line", Journal of Physics: Conference Series **1992** (2021) 032067. https://doi.org/10.1088/1742-6596/1992/3/032067.

[7] M. Park & J. Jeong, "Design and implementation of machine vision-based quality inspection system in mask manufacturing process", Sustainability **14** (2022) 6009. https://doi.org/10.3390/su14106009.

[8] C. A. de Souza Silva & E. P. Paladini, "Smart machine vision system to improve decision-making on the assembly line", Machines **13** (2025) 98. https://doi.org/10.3390/machines13020098.

[9] S. J. More & A. J. More, "Computer vision & IoT based smart conveyor belt sorting system", International Journal of Engineering Research & Technology **13** (2024) IJERTV13IS060066. Available online: https://www.ijert.org/computer-vision-iot-based-smart-conveyor-belt-sorting-system.

[10] I. Yousif, L. Burns, F. El Kalach & R. Harik, "Leveraging computer vision towards high-efficiency autonomous industrial facilities", Journal of Intelligent Manufacturing **36** (2025) 2983. https://doi.org/10.1007/s10845-024-02396-1.

[11] J.-S. Chou & C.-H. Liu, "Automated sensing system for real-time recognition of trucks in river dredging areas using computer vision and convolutional deep learning", Sensors **21** (2021) 555. https://doi.org/10.3390/s21020555.

[12] S. R. Addula & A. K. Tyagi, "Future of computer vision and industrial robotics in smart manufacturing", in Artificial Intelligence-Enabled Digital Twin for Smart Manufacturing, A. K. Tyagi, S. Tiwari, S. K. Arumugam & A. K. Sharma (Eds.), Wiley, Hoboken, NJ, USA, 2024, pp. 505–539. https://doi.org/10.1002/9781394303601.ch22.

[13] R. Bhana, H. Mahmoud & M. Idrissi, "Smart industrial safety using computer vision", 2023 28th International Conference on Automation and Computing, Birmingham, UK, 2023, pp. 1–6. https://doi.org/10.1109/ICAC57885.2023.10275164.

[14] J. L. Mira, J. Barba, F. P. Romero, M. S. Escolar, J. Caba & J. C. López, "Benchmarking of computer vision methods for energy-efficient high-accuracy olive fly detection on edge devices", Multimedia Tools and Applications **83** (2024) 81785. https://doi.org/10.1007/s11042-024-18589-y.

[15] G. Li, Z. Liu, Z. Feng, J. Lyu, B. Li, G. Chen & Q. Yao, "Research on a machine vision-based electro-killing pheromone-baited intelligent agricultural pest monitoring method", Frontiers in Plant Science **16** (2025) 1521594. https://doi.org/10.3389/fpls.2025.1521594.

[16] H. A. Zhou, D. Wolfschläger, C. Florides, J. Werheid, H. Behnen, J.-H. Woltersmann, T. C. Pinto, M. Kemmerling, A. Abdelrazeq & R. H. Schmitt, "Generative AI in industrial machine vision: a review", Journal of Intelligent Manufacturing **37** (2026) 1447. https://doi.org/10.1007/s10845-025-02604-6.

[17] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting & D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", BMJ **372** (2021) n71. https://doi.org/10.1136/bmj.n71.

[18] A. Shaushenova, O. Kuznetsov, A. Nurpeisova & M. Ongarbayeva, "Implementation of Kolmogorov–Arnold networks for efficient image processing in resource-constrained Internet of Things devices", Technologies **13** (2025) 155. https://doi.org/10.3390/technologies13040155.

[19] R. Rani, S. Kumar, S. K. Pippal, M. Gund, U. Chaudhari, R. Agrawal, M. Dalsaniya & L. Verma, "IPS: intelligent parking system using YOLO and image processing", International Journal of Transport Development and Integration **8** (2024) 447. https://doi.org/10.18280/ijtdi.080308.

[20] J. Shreyas, A. P. Jyothi & H. M. Mallaradhya, "Smart robotic surgical assistant using voice command and image processing", International Journal of Multidisciplinary Research and Analysis **7** (2024) 5457. https://doi.org/10.47191/ijmra/v7-i12-12.

[21] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan & M. Shah, "Transformers in vision: a survey", ACM Computing Surveys **54** (2022) 200. https://doi.org/10.1145/3505244.

[22] S. Pak, S. G. Park, J. Park, H. R. Choi, J. H. Lee, W. Lee, S. T. Cho, Y. G. Lee & H. Ahn, "Application of deep learning for semantic segmentation in robotic prostatectomy: comparison of convolutional neural networks and visual transformers", Investigative and Clinical Urology **65** (2024) 551. https://doi.org/10.4111/icu.20240159.

[23] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger & I. Sutskever, "Learning transferable visual models from natural language supervision", Proceedings of Machine Learning Research **139** (2021) 8748. Available online: https://proceedings.mlr.press/v139/radford21a.html.

[24] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár & R. Girshick, "Segment Anything", 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2023, pp. 4015–4026. Available online: https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html.

[25] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski & A. Joulin, "Emerging properties in self-supervised vision transformers", 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021, pp. 9650–9660. Available online: https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper.html.