Journal of Scientific Insights
E-ISSN: 3062-8571
DOI: https://doi.org/10.69930/jsi.v3i4.826
Review Article

ScienceTech Group

Vol. 3 (4), 2026
Page: 423-436

Systematic Review of Computer Vision Performance for Age-Friendly Healthcare Facility Auditing: A Task-Based Evidence Synthesis

Raden Roro Sonya Dewi Wulandari [1,*], Ramadhan Sanyoto S. Widodo [2], Muhammad Febriano S. Suwarto [1], Tony Sugiarso [3], Ahdun Trigono [1]

[1] Hospital Administration Postgraduate Program, Faculty of Health Sciences, Universitas Respati Indonesia, Indonesia
[2] Information Systems Program Study, Faculty of Information Technology, Universitas Respati Indonesia, Indonesia
[3] Computer Science Study Program, Faculty of Information Technology, Universitas Respati Indonesia, Indonesia
*Email (corresponding author): sonya@urindo.ac.id

Abstract. *Population aging increases the need for healthcare environments that are safe, accessible, and supportive of mobility for older adults. Facility auditing remains an important part of quality improvement, but manual inspection depends on staff time, checklist interpretation, and the moment of observation. This systematic review synthesizes task-based evidence on computer vision for age-friendly healthcare facility auditing. The review followed PRISMA 2020 and PRISMA-S principles. Eligible studies used image, video, RGB-D, depth, infrared, or thermal computer vision in hospitals, intensive care units, wards, nursing homes, long-term care, assisted living, or closely related care environments. Studies were included if they reported task-level performance or demonstrated clear operational relevance to environmental safety, accessibility, hygiene, mobility, posture, falls, or indoor object recognition. The evidence was synthesized narratively because the studies differed widely in their settings, sensor types, target labels, datasets, and metrics. The evidence clustered into environment understanding, hygiene and contact monitoring, mobility and activity monitoring, and posture or fall-related analysis. Technical performance was often promising within local datasets, but few studies translated model outputs into audit indicators, facility dashboards, or administrative decisions. Future work should align computer vision outputs with age-friendly audit tools, validate models across facilities, report privacy, workflow fit, and operational usefulness.*

Keywords: *Computer Vision; Age-Friendly Healthcare; Facility Auditing; Healthcare Environment*

1. Introduction

Population aging changes how healthcare organizations manage buildings, services, and digital transformation. Older adults are more likely to experience frailty, multimorbidity, falls, delirium, and functional decline when care environments do not support safe movement and clear orientation (Reeves et al., 2018; Montero-Odasso et al., 2022; Wu et al., 2023). The problem is therefore not only clinical because rooms, corridors, beds, chairs, signage, and staff workflows also shape whether older patients can move safely during care (WHO, 2007a; IHI, 2020).

The Age-Friendly Health Systems movement uses the 4Ms framework: What Matters, Medication, Mentation, and Mobility (IHI, 2020; Fulmer et al., 2018). Mobility is closely connected to the built environment because safe movement depends on uncluttered

Received: 25 June 2026 / Accepted: 21 July 2026 / Available online: 24 July 2026

pathways, visible assistive devices, appropriate beds and chairs, reachable handrails, readable signage, and staff workflows that reduce the risk of falls and infections (IHI, 2020; WHO, 2007b). Age-friendly facility auditing should therefore be treated as part of safety management and quality improvement rather than a decorative checklist (Hut-Mossel et al., 2021; Victorian Department of Health, 2015).

Facility audits are usually performed through manual observation, paper-based checklists, and periodic inspection by administrators or quality-assurance teams (Hut-Mossel et al., 2021). This approach is understandable because many environmental features still require human judgment and local context (Victorian Department of Health, 2015). However, manual auditing is not easily scalable because it depends on auditor availability, inspection timing, and staff ability to notice dynamic risks such as blocked walkways, misplaced equipment, or missed hand-hygiene opportunities (Hut-Mossel et al., 2021; WHO, 2021).

Computer vision can become a digital support layer for this audit process. Modern image and video algorithms can classify scenes, detect objects, segment walkable areas, estimate human posture, recognize activities, and track events from RGB cameras, depth sensors, RGB-D systems, infrared cameras, or thermal sensors (Krizhevsky et al., 2012; He et al., 2016; Cao et al., 2017). These functions are relevant to facility auditing. For instance, preliminary studies can use computer vision for tracking and hand-hygiene monitoring in hospital spaces (Haque et al., 2017), developed a deep-learning system to detect patient mobilization in the ICU (Yeung et al., 2019), and evaluated automatic hand-hygiene detection using computer vision (Singh et al., 2020).

In healthcare and elder-care environments, computer vision has already been explored for hand hygiene monitoring, ICU patient mobilization, indoor object recognition, nursing-home object classification, posture monitoring, and fall detection. These studies are not always labeled as age-friendly research, but many measure audit-relevant targets such as hygiene behavior, mobility events, room objects, staff-patient contact, and fall-related posture (Igual et al., 2013; Mubashir et al., 2013; Delahoz & Labrador, 2014). The challenge is that technical performance does not automatically become administrative evidence unless outputs are mapped to audit indicators and management actions (Sendak et al., 2020; Vasey et al., 2022).

The evidence base is fragmented because studies use different sensors, tasks, labels, datasets, and metrics. A hand-hygiene system may report sensitivity and specificity, while an object-detection dataset may report mean average precision, a segmentation model may report Dice or intersection-over-union, and an activity-recognition system may report event-level accuracy (Padilla et al., 2020; Reinke et al., 2024). These metrics address different questions, so this review does not combine them into a single pooled estimate (Page et al., 2021b; Whiting et al., 2011).

This review aims to synthesize computer vision performance for age-friendly healthcare facility auditing and to determine whether model outputs are linked to audit indicators or administrative workflows. The review asks what computer vision tasks have been studied in healthcare or elder-care environments, what performance evidence has been reported, and how far those outputs can support facility auditing for older adults. The contribution of this paper is a task-based evidence synthesis rather than a meta-analysis, because the literature is currently too heterogeneous for a single quantitative summary. This review aims to synthesize computer vision performance for age-friendly healthcare facility auditing, thereby supporting Sustainable Development Goal 3 of ensuring healthy lives and promoting well-being for all at all ages.

2. Methods

This review followed the PRISMA 2020 reporting guideline and the PRISMA-S extension for literature search reporting (Page et al., 2021a; Rethlefsen et al., 2021). A protocol was prepared using an Open Science Framework-style structure that covered objectives, eligibility criteria, information sources, screening, data extraction, quality appraisal, and synthesis plan (Page et al., 2021b). OSF-style registration was selected because this review focuses on technology performance and audit relevance rather than direct treatment effects in patients.

The review question was organized using a modified PICOS/T framework (Page et al., 2021a). The population and setting included hospitals, ICUs, wards, rehabilitation units, nursing homes, long-term care facilities, assisted living facilities, and closely related elder-care environments. The index technology was image- or video-based computer vision, including RGB, RGB-D, depth, infrared, and thermal sensing. The target constructs were environmental safety, accessibility, mobility support, hygiene-monitoring opportunities, patient activity, staff-patient contact near risk points, indoor object presence, room or scene type, posture, falls, and fall risk.

The search strategy combined three conceptual blocks: computer vision technology, healthcare or elder-care setting, and audit-relevant targets. The technology block included terms such as "computer vision", "deep learning", "object detection", "scene classification", "activity recognition", "pose estimation", "semantic segmentation", "RGB-D", and "depth camera". The setting block included "hospital", "ICU", "ward", "nursing home", "care home", "long-term care", and "assisted living". The audit target block included "age-friendly", "older adult", "elderly", "mobility", "fall", "accessibility", "handrail", "wheelchair", "hand hygiene", "signage", "obstacle", and "bedside monitoring".

Records were searched in Scopus, PubMed, IEEE Xplore, Web of Science, and supplementary citation searching from 2010 to 2025. The lower year limit was selected because deep learning, depth sensing, and RGB-D systems became more visible in computer vision research around this period (Krizhevsky et al., 2012; Silberman et al., 2012; Shotton et al., 2011). The search was limited to English-language publications for feasibility.

All reviewers independently screened titles and abstracts before assessing full texts against the eligibility criteria. Disagreements were resolved through discussion, which is recommended to reduce selection bias in systematic reviews (Page et al., 2021b; Whiting et al., 2011). The selection process was documented using a PRISMA-style flow diagram to ensure transparency in identification, screening, eligibility, and inclusion (Page et al., 2021a).

Data extraction captured the publication year, country, setting, sensor modality, task family, dataset source, annotation method, model architecture, validation strategy, performance metrics, privacy measures, and whether the system was linked to an audit indicator, a dashboard, or an administrative workflow. These variables were chosen because computer vision performance depends not only on the model but also on dataset quality, annotation quality, validation design, and deployment context (Moons et al., 2015; Reinke et al., 2024; Sendak et al., 2020).

Study quality was assessed using a tailored tool combining QUADAS-2 items for bias and applicability, CLAIM items for AI reporting completeness, and PROBAST items for studies that predicted fall risk or patient-level outcomes (Whiting et al., 2011; Mongan et al., 2020; Wolff et al., 2019). The quality domains included dataset selection, reference standard, annotation quality, train-validation-test split, external validation, threshold choice, privacy protection, and workflow reporting (Moons et al., 2015; Vasey et al., 2022; Reinke et al., 2024).

No meta-analysis was conducted because the included studies differed strongly in settings, sensors, labels, tasks, and outcome metrics. Narrative synthesis grouped the studies into four task families: environment understanding; hygiene and contact monitoring; mobility and activity monitoring; and posture or fall-related analysis. This approach was more appropriate than a pooled estimate because the review question concerns audit relevance across heterogeneous computer vision tasks. Table 1 presents the PICOS/T framework used to restrict the review to audit-relevant computer vision rather than the broader field of medical image diagnosis.

Table 1. PICOS/T eligibility framework for the review

Element	Inclusion Criteria	Exclusion Criteria
Population/setting	Hospitals, ICUs, wards, rehabilitation units, nursing homes, long-term care, assisted living, and elder-care environments.	Pure domestic smart-home studies without healthcare or elder-care relevance.
Index technology	RGB, RGB-D, depth, infrared, thermal, image, or video computer vision.	Non-visual sensors only, administrative databases only, or medical diagnosis from radiology/pathology images.
Target construct	Accessibility, mobility, hygiene, room/scene type, indoor objects, posture, falls, contact, or safety events.	Pure diagnostic prediction without environmental or operational audit relevance.
Outcome	Task-level performance metrics and/or audit relevance.	No reportable performance data or only opinion-based commentary.
Synthesis type	Narrative task-based synthesis across comparable task families.	Cross-task quantitative pooling or unsupported meta-analysis.

3. Results and Discussion

The search identified 714 records from databases and supplementary citation searching. After 183 duplicate records were removed, 531 records were screened by title and abstract. At this stage, 438 records were excluded because they did not meet the setting, technology, or audit-relevance criteria. Ninety-three full-text reports were sought, eight reports could not be retrieved, and 85 full-text reports were assessed for eligibility. Sixty-two full-text reports were excluded because they were outside healthcare or elder-care settings, did not use computer vision as the main method, focused on clinical diagnosis only, lacked audit relevance, lacked performance data, or were duplicate or secondary reports. Twenty-three studies were included in the qualitative task-based evidence synthesis. Overall process of summarising the study selection process as illustrated by Figure 1.

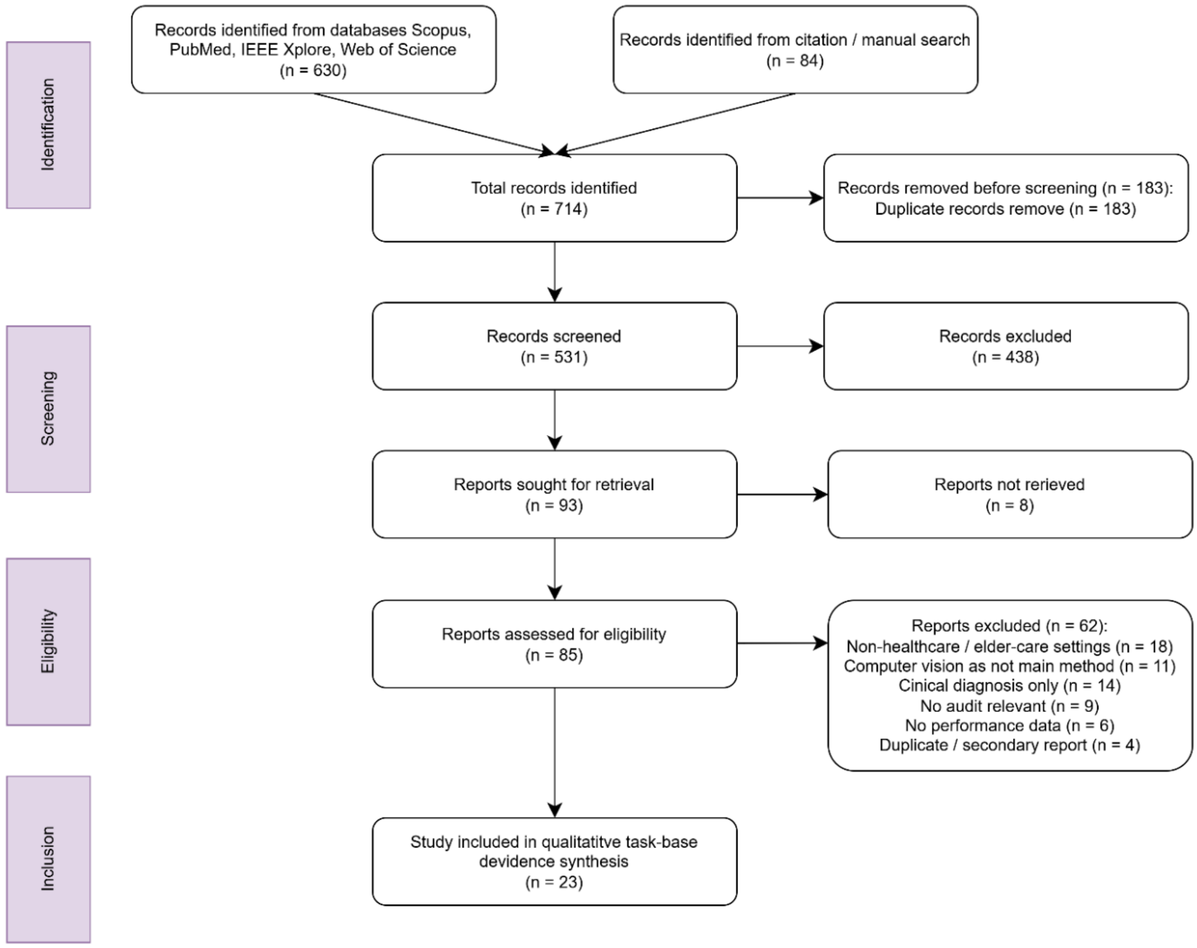

Figure 1. PRISMA flow diagram summarising the study selection process

The included studies were published between 2010 and 2025 and covered hospitals, ICUs, wards, nursing-home settings, long-term care, rehabilitation, and simulated care environments (Haque et al., 2017; Singh et al., 2020; Yeung et al., 2019; Ismail et al., 2020). Data modalities included RGB images, video streams, depth cameras, RGB-D sensors, infrared imaging, and thermal sensing (Silberman et al., 2012; Shotton et al., 2011; Cao et al., 2017). The main tasks for families were environmental understanding, hygiene and contact monitoring, mobility and activity monitoring, and posture- or fall-related analysis. Table 2 lists anchor studies that represent the main task families and shows that technical performance must be interpreted together with audit relevance.

Table 2. Anchor studies and audit relevance

Study	Task	Computer Vision Target	Reported Performance or Contribution	Administrative Relevance
Haque et al. (2017)	Vision-based smart hospital monitoring.	Hospital setting with staff/patient contact and hand hygiene zones.	Tracking and spatial analytics for contact and hygiene monitoring.	Supports infection-prevention observation, but dashboard translation remains limited.

Study	Task	Computer Vision Target	Reported Performance or Contribution	Administrative Relevance
Singh et al. (2020)	Automatic hand hygiene detection.	Hospital depth camera data around dispensers.	High sensitivity and specificity for dispenser-use detection.	Relevant to the audit of infection-prevention behavior in older-adult care.
Yeung et al. (2019)	Patient mobilization activity detection.	ICU setting with computer vision for bed and chair movement.	Deep learning-based recognition of mobilization events.	Relevant to Mobility in the 4Ms framework.
Ismail et al. (2020)	Nursing-home indoor object classification.	MYNursingHome dataset with aged-care object categories.	CNN-based classification of indoor care objects.	Supports object inventory and facility readiness in elder-care settings.
Hu et al. (2023)	Hospital indoor object detection.	Hospital indoor object dataset with facility objects.	Object detection benchmarking using modern detectors.	Supports environmental perception but needs audit mapping.
Igual et al. (2013); Mubashir et al. (2013)	Vision-based fall and posture monitoring.	Care or smart-monitoring environments with camera-based fall signals.	Classification of posture, falling, or abnormal movement.	Relevant to safety, but facility causes must be examined.

3.1. Environment Understanding

Environment understanding includes room classification, indoor object detection, and free-space or obstacle analysis. These tasks are relevant because age-friendly audits need to know whether spaces are ready for older users, whether key mobility aids are available, and whether pathways are blocked (WHO, 2007b; Victorian Department of Health, 2015).

Object detection models such as Faster R-CNN, SSD, Mask R-CNN, YOLO, DETR, and Swin Transformer provide the technical basis for detecting beds, chairs, wheelchairs, medical equipment, signs, and other indoor objects (Ren et al., 2015; Liu et al., 2016; He et al., 2017; Bochkovskiy et al., 2020; Carion et al., 2020; Liu et al., 2021). However, models trained on general datasets such as ImageNet, COCO, Places, ADE20K, or ScanNet may not capture care-specific objects and audit features without domain adaptation (Deng et al., 2009; Lin et al., 2014; Zhou et al., 2017; Zhou et al., 2019; Dai et al., 2017).

Domain-specific datasets are more directly useful for this review. The MYNursingHome dataset is valuable because it moves indoor object recognition closer to aged-care settings (Ismail et al., 2020). Moreover, hospital indoor object detection datasets are also useful, but many still underrepresent audit targets such as ramp condition, handrail availability, wheelchair turning space, bedside clearance, and signage readability (Hu et al., 2023).

3.2. Hygiene and Contact Monitoring

Hygiene and contact monitoring are important because infection prevention is a core patient-safety issue, particularly for older adults with frailty or multimorbidity (WHO, 2021; Wu et al., 2023). Vision-based hand hygiene systems have used cameras or depth sensors to detect dispenser use, staff movement, and contacts near patient zones (Haque et al., 2017; Singh et al., 2020). Singh et al. (2020) reported strong sensitivity and specificity for hand hygiene detection using computer vision (Singh et al., 2020). Haque et al. (2017) showed that tracking and spatial analytics can support hand hygiene compliance monitoring in smart-hospital settings (Haque et al., 2017). These studies show audit potential but often stop at event detection and do not always explain how outputs become staff feedback, infection-control indicators, or quality-improvement actions (Hut-Mossel et al., 2021; Sendak et al., 2020).

3.3. Mobility and Activity Monitoring

Mobility and activity monitoring includes bed-to-chair transfers, ambulation, patient mobilization, staff-patient interactions, and periods of low movement. These tasks are directly connected to the 4Ms Mobility domain because older adults need environments that promote safe mobility rather than unnecessary bed rest (IHI, 2020; Montero-Odasso et al., 2022). Yeung et al. (2019) demonstrated that deep learning-based computer vision can detect patient mobilization activities in ICU settings. Pose estimation and skeleton-based action recognition methods provide additional technical foundations for recognizing transfers, body movements, and posture over time (Shotton et al., 2011; Cao et al., 2017; Yan et al., 2018). The limitation is that event-level recognition does not automatically assess whether the room layout, assistive device placement, clutter, lighting, or staff workflow supports safe mobility. For age-friendly auditing, the model output must be connected to questions about readiness for safe movement, transfer, recovery, and independence (IHI, 2020; Fulmer et al., 2018; Victorian Department of Health, 2015).

3.4. Posture and Fall-Risk Analysis

Fall and posture analysis is a visible area where computer vision intersects with age-friendly care. Reviews of fall detection systems show that vision-based systems can detect lying, standing, sitting, falling, and abnormal movement, but their deployment in care settings raises concerns about privacy, camera placement, and false alarms (Igual et al., 2013; Mubashir et al., 2013; Delahoz & Labrador, 2014). Segmentation and medical image architectures, such as fully convolutional networks and U-Net, are relevant technical references for pixel-level analysis, although their original use is broader than facility auditing (Long et al., 2015; Ronneberger et al., 2015). In facility auditing, fall detection should be linked to environmental factors such as lighting, obstacles, bed height, handrail placement, and response workflow, rather than treated solely as an alert classification problem (Montero-Odasso et al., 2022; WHO, 2007b).

3.5. From Model Output to Audit Action

Figure 2 summarises the conceptual bridge required for practical use. Visual inputs feed the computer vision task layer, which produces observable audit features; those features must then be translated into audit indicators and management actions. The framework shows why technical accuracy alone is not enough (Mongan et al., 2020; Reinke et al., 2024). A bounding box around a wheelchair becomes useful only when it answers an audit question, such as whether mobility aids are available, whether a pathway is blocked, or whether a transfer area is ready for safe use (Victorian Department of Health, 2015; IHI, 2020). Similarly, a detected hand-hygiene event is useful only when it can be interpreted within a care opportunity, a staff workflow, and an infection-control feedback cycle (Singh et al., 2020; WHO, 2021). Therefore, Table 3 maps computer vision outputs to audit indicators because a useful audit system must convert into decisions that administrators, nurses, managers, and healthcare worker teams can understand.

Table 3. Mapping computer vision outputs to age-friendly audit indicators

Computer Vision Output	Potential Audit Indicator	Administrative Use
Room classification (Zhou et al., 2017).	Room type and care-zone identity.	Check room readiness and match the audit checklist by area.
Object detection of beds, chairs, wheelchairs, and dispensers (Hu et al., 2023; Ren et al., 2015).	Availability and position of key equipment.	Prioritize equipment placement and remove obstacles.
Free-space or obstacle segmentation (Long et al., 2015; Ronneberger et al., 2015).	Clear pathway and walking-space compliance.	Trigger maintenance or layout correction.
Hand hygiene event detection (Haque et al., 2017; Singh et al., 2020).	Infection-prevention observation.	Support staff feedback and quality-improvement cycles.
Patient mobilisation detection (Yeung et al., 2019; Yan et al., 2018).	Safe mobility support and transfer readiness.	Identify low-mobility periods and care-flow barriers.
Posture or fall-risk signals (Igual et al., 2013; Mubashir et al., 2013).	Fall-risk flag and unsafe movement pattern.	Coordinate human review, response workflow, and environmental correction.

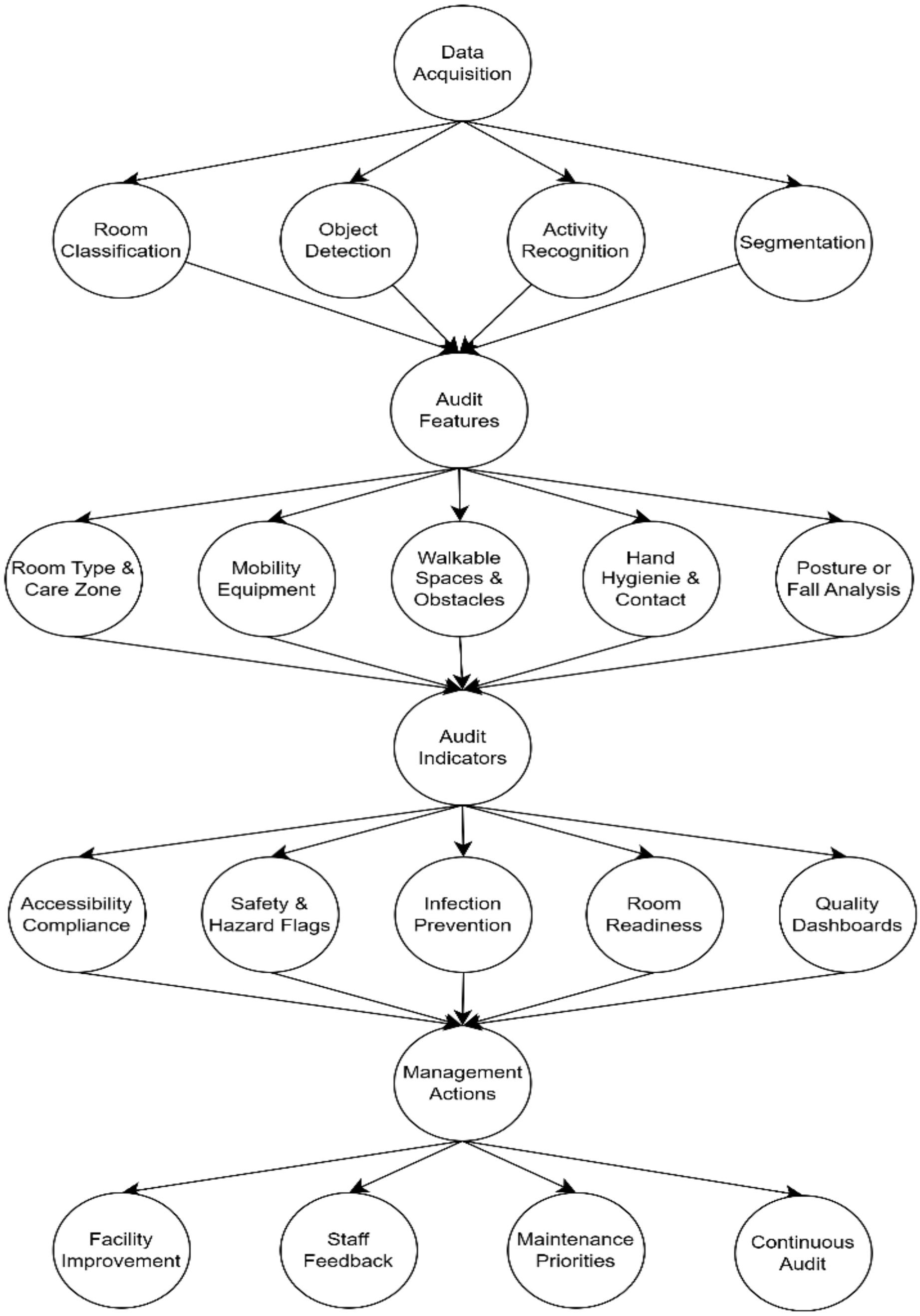

Figure 2. Conceptual framework linking computer vision tasks to age-friendly healthcare facility auditing and management action

Overall, the strongest evidence is for task feasibility rather than administrative effectiveness. Hand-hygiene detection and ICU mobilization detection demonstrate that computer vision can recognize selected events in care environments, but by themselves, they do not prove improvements in facility readiness or older-adult safety (Haque et al., 2017; Yeung et al., 2019; Singh et al., 2020). The main evidence gap is implementation. Many papers report accuracy, sensitivity, specificity, mAP, Dice, and other technical metrics, but fewer explain how these outputs are used by nurses, administrators, infection-control teams, or facility managers. Reporting guidelines for AI evaluation and clinical implementation emphasize that workflow, user interaction, data handling, and deployment context should be described, not only model performance (Mongan et al., 2020; Vasey et al., 2022; DECIDE-AI Steering Group, 2022).

There is also a dataset gap. General computer-vision datasets are useful for pre-training and benchmarking, but they are not designed around age-friendly audit items such as handrail reachability, wheelchair turning area, obstacle-free transfer space, signage readability, and bedside clearance. Future datasets should explicitly include these labels and should test performance across different facilities, lighting conditions, camera placements, and patient populations (Deng et al., 2009; Lin et al., 2014; Zhou et al., 2019; Reinke et al., 2024).

Finally, the use of cameras in healthcare must be handled carefully. Computer vision systems may improve observation, but they may also raise privacy, consent, surveillance, and bias concerns if deployed without adequate governance. For this reason, edge processing, de-identification, limited retention, clear accountability, and user involvement should be considered in future studies (Sendak et al., 2020; Vasey et al., 2022). Researchers should work with hospital administrators, nurses, older adults, and facility managers to define audit indicators before training models. Without that collaboration, computer vision may remain a technical demonstration rather than a tool that improves age-friendly facility readiness.

Conclusions

This systematic review indicates that computer vision has clear potential to support audits of age-friendly healthcare facilities, particularly for environmental assessment, hygiene and contact monitoring, mobility and activity monitoring, and posture or fall analysis. However, the existing evidence is stronger for technical feasibility than for administrative implementation or for demonstrated improvements in older adults' safety. The most important gap is not only in algorithm accuracy but also in translating into audit indicators, dashboards, staff feedback, and facility improvements. Future work should build age-friendly audit datasets, validate models across multiple facilities, and report privacy, workflow fit, and operational usefulness. If these conditions are addressed, computer vision can become a useful support tool for safer, more accessible, and more inclusive care environments for older adults.

Funding

This project was supported by the Respati Education Board. The funder had no role in the study design, data collection, data analysis, interpretation of findings, or manuscript preparation.

Conflicts of Interest

The authors declare no conflict of interest.

References

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., & Savarese, S. (2016). 3d semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1534-1543).

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Cao, Z., Simon, T., Wei, S. E., & Sheik, Y. (2017). Real-time multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In European Conference on Computer Vision (pp. 213-229). Cham: Springer International Publishing.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., et al. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine, 162(1), W1-W73.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., & Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5828-5839).

DECIDE-AI Steering Group. (2022). DECIDE-AI: New reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nature Medicine, 28, 186-187.

Delahoz, Y. S., & Labrador, M. A. (2014). Survey on fall detection and fall prevention using wearable and external sensors. Sensors, 14(10), 19806-19842.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009, the IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255). IEEE.

Fulmer, T., Mate, K. S., & Berman, A. (2018). The age-friendly health system imperative. Journal of the American Geriatrics Society, 66(1), 22-24.

Haque, A., Guo, M., Alahi, A., Yeung, S., Luo, Z., Rege, A., et al. (2017). Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. In Machine Learning for Healthcare Conference (pp. 75-87). PMLR.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2961-2969).

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

Hu, D., Li, S., & Wang, M. (2023). Object detection in hospital facilities: A comprehensive dataset and performance evaluation. Engineering Applications of Artificial Intelligence, 123, 106223.

Hut-Mossel, L., Ahaus, K., Welker, G., & Gans, R. (2021). Understanding how and why audits work in improving the quality of hospital care: a systematic realist review. PloS one, 16(3), e0248677.

Igual, R., Medrano, C., & Plaza, I. (2013). Challenges, issues, and trends in fall detection systems. Biomedical engineering online, 12(1), 66.

Institute for Healthcare Improvement. (2020). Age-friendly health systems: Guide to using the 4Ms in the care of older adults. Institute for Healthcare Improvement.

Ismail, A., Ahmad, S. A., Soh, A. C., Hassan, M. K., & Harith, H. H. (2020). MYNursingHome: A fully-labeled image dataset for indoor object classification. Data in Brief, 32, 106268.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft Coco: Common objects in context. In European Conference on Computer Vision (pp. 740-755). Cham: Springer International Publishing.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single-shot multibox detector. In the European Conference on Computer Vision (pp. 21-37). Cham: Springer International Publishing.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE International Conference on Computer Vision.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Mongan, J., Moy, L., & Kahn, C. E. (2020). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. Radiology: Artificial Intelligence, 2(2), e200029.

Montero-Odasso, M., Van Der Velde, N., Martin, F. C., Petrovic, M., Tan, M. P., Ryg, J., et al. (2022). World guidelines for falls prevention and management for older adults: a global initiative. Age and aging, 51(9), afac205.

Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. Neurocomputing, 100, 144-152.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021a). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021b). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ, 372.

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. In the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 237-242). IEEE.

Reeves, D., Pye, S., Ashcroft, D. M., Clegg, A., Kontopantelis, E., Blakeman, T., & van Marwijk, H. (2018). The challenge of aging populations and patient frailty: can primary care adapt?. Bmj, 362.

Reinke, A., Tizabi, M. D., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Kavur, A. E., et al. (2024). Understanding metric-related pitfalls in image analysis validation. Nature Methods, 21(2), 182-194.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28.

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Systematic reviews, 10(1), 39.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Cham: Springer International Publishing.

Sendak, M., Elish, M. C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., et al. (2020). " The human body is a black box," supporting clinical decision-making with deep learning. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 99-109).

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., et al. (2011). Real-time human pose recognition in parts from single depth images. In CVPR 2011 (pp. 1297-1304). IEEE.

Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision (pp. 746-760). Berlin, Heidelberg: Springer Berlin Heidelberg.

Singh, A., Haque, A., Alahi, A., Yeung, S., Guo, M., Glassman, J. R., et al. (2020). Automatic detection of hand hygiene using computer vision technology. Journal of the American Medical Informatics Association, 27(8), 1316-1320.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339.

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., et al. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ, 377.

Victorian Department of Health. (2015). Hospital Environment Audit Tool (HEAT). Victorian Government Department of Health.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine, 155(8), 529-536.

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Annals of internal medicine, 170(1), 51-58.

World Health Organization. (2007a). Global age-friendly cities: A guide. World Health Organization.

World Health Organization. (2007b). Age-friendly primary health care centers toolkit. World Health Organization.

World Health Organization. (2021). Global patient safety action plan 2021-2030: towards eliminating avoidable harm in health care. World Health Organization.

Wu, J., Zhang, H., Shao, J., Chen, D., Xue, E., Huang, S., et al. (2023). Healthcare for older adults with multimorbidity: a scoping review of reviews. Clinical Interventions in Aging, 1723-1735.

Yan, S., Xiong, Y., & Lin, D. (2018). Spatial-temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

Yeung, S., Rinaldo, F., Jopling, J., Liu, B., Mehra, R., Downing, N. L., et al. (2019). A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. NPJ digital medicine, 2(1), 11.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452-1464.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic understanding of scenes through the ade20k dataset. International journal of computer vision, 127(3), 302-321.