DOI: 10.66069/ojspub.27450843

Artificial Intelligence–Enabled Quality Control in Diagnostic Imaging: Evidence, Evaluation Requirements, and Human-Supervised Governance

Zeming He, Tian Zhu*, Xuan Luo, Yue Qiu, Mi Zhang, Xuejie Xu, Qiuting Guo

Xianyang Polytechnic Institute, Xianyang 712000, Shaanxi, China
**Correspondence Author*

Abstract: ***Objective:*** ***To review AI-enabled quality control (QC) across diagnostic imaging modalities, distinguish algorithmic performance from transportability and clinical effectiveness, and propose an evaluation and governance framework for human-supervised QC agents.*** ***Methods:*** ***PubMed was searched in a targeted structured manner for English-language publications from January 2017 through August 4, 2026. Peer-reviewed studies were prioritized when they directly assessed acquisition quality, positioning and coverage, artifacts, protocol and dose, reconstruction, or equipment performance. Relevant professional statements were also traced. Findings were synthesized narratively, with SANRA items used for methodological self-checking; no meta-analysis or formal risk-of-bias assessment was performed.*** ***Results:*** ***Published applications map to five functions: examination assurance, acquisition assurance, output assurance, system assurance, and learning and governance. Evidence in radiography, CT, and MRI is dominated by retrospective, task-specific technical validation. Ultrasound and mammography include observational workflow studies after deployment. Patient-image evidence in nuclear medicine remains comparatively sparse, whereas phantom and equipment surveillance are closer to routine implementation. Recurrent limitations include inconsistent quality definitions, subjective labels, inadequate handling of leakage and clustering, limited external validation, and weak links between technical scores and diagnostic utility.*** ***Conclusion:*** ***AI can broaden QC coverage and accelerate feedback, but a universal technical score should not replace professional judgment. Human-supervised QC agents should coordinate specialized models and deterministic rules and should be constrained by human authorization, stage-gated validation, and life-cycle monitoring.***

Keywords: Artificial intelligence, Medical imaging, Quality control, Quality assurance, Human–AI collaboration, External validation, Life-cycle governance**.**

1. Introduction

Medical image quality is not merely an aesthetic preference. It is a composite property reflecting whether anatomical coverage, acquisition geometry, artifacts, contrast, and noise are sufficient for a specific diagnostic task. In routine practice, radiographers and sonographers assess acquisition quality at the console, interpreting physicians identify deficiencies during review, medical physicists perform scheduled equipment testing, and managers audit repeat or recall performance. These safeguards are indispensable, but their coverage may be incomplete, their feedback delayed, and their outputs difficult to aggregate continuously.

When integrated with imaging workflows, AI can in principle analyze examinations in real or near-real time and summarize recurring deficiencies by device, protocol, operator, or period. However, the label *AI quality control* is applied to tasks at fundamentally different levels, including chest-radiograph rotation, CT scan range, MRI motion, ultrasound standard planes, mammographic positioning, and phantom-image assessment. Their reference standards, failure consequences, and opportunities for correction differ; they cannot be reduced to one universal quality score.

This review addresses a common clinical question: Can AI detect, explain, or prevent a correctable loss of imaging quality while useful action remains possible? It compares evidence across modalities and distinguishes algorithmic performance from workflow effectiveness and clinical utility. Appraisal was informed by CLAIM 2024, DECIDE-AI, STARD-AI, and multisociety guidance for imaging AI [1–4]. These documents are reporting or implementation frameworks, however, and should not be misrepresented as formal instruments for grading certainty of evidence.

2. Review Methods

2.1 Scope and Definitions

This was a targeted structured narrative review. SANRA items were used to self-check the justification of the review, description of the search, referencing, scientific reasoning, and presentation of conclusions [5]. Here, *quality control* primarily denotes detection and correction at the level of an examination, series, image, or device output, whereas *quality assurance* denotes the broader processes, training, equipment oversight, audit, and governance required for sustained performance. Both are incorporated into one closed-loop framework for cross-modality comparison, but equipment physics QC and patient-image quality assessment are treated as distinct tasks.

The scope was limited to diagnostic radiography, CT, MRI, ultrasound, mammography, and PET/SPECT. Pathology, ophthalmic imaging, radiotherapy imaging, interventional navigation, and studies limited to image generation or enhancement without direct QC evaluation were outside the core scope. Accordingly, *diagnostic imaging modalities* does not denote multimodal data fusion or a general-purpose

multimodal foundation model.

2.2 Search and Eligibility Approach

PubMed was searched for English-language literature published from January 1, 2017, through August 4, 2026; the search was last run on August 4, 2026. Search concepts included *artificial intelligence, machine learning, quality control, quality assurance, image quality, positioning, standard plane, artifact*, and *scan range*, combined separately with radiography, CT, MRI, ultrasound, mammography, PET, SPECT, and nuclear medicine. The reproducible search string is provided in Appendix 1. Reference lists of eligible papers and professional statements were also traced.

Priority was given to peer-reviewed studies that directly evaluated technical quality, anatomical coverage, positioning, artifacts, protocol and dose, reconstruction, or equipment performance in patient or phantom images. Pure diagnostic or image-enhancement studies were not treated as core evidence unless they directly evaluated acquisition quality or a QC mechanism.

2.3 Narrative Synthesis

Because quality definitions, units of analysis, reference standards, and outcomes differed substantially, studies were compared narratively by modality and QC function. Candidate reports were first judged for direct QC relevance from titles and abstracts; priority was then given to representative studies with clearer reference standards, stronger validation designs, or workflow outcomes. This was not exhaustive duplicate screening. Table 1 summarizes representative directly relevant study designs and the principal evidence gaps. It does not combine sample size, risk of bias, or effect estimates and should not be interpreted as GRADE or another formal certainty-of-evidence rating.

3. Functional Framework for AI-Enabled Imaging QC

AI-enabled QC can be organized into five connected functions (Figure 1). Examination assurance verifies the order, patient, body region, laterality or orientation, protocol, and series. Acquisition assurance evaluates centering, positioning, anatomical coverage, standard planes, motion, and exposure. Output assurance identifies artifacts and assesses noise, sharpness, contrast, and task-specific diagnostic adequacy. System assurance monitors dose, repeats, reconstruction, phantom images, and device or model drift. Learning and governance convert outputs into operator feedback, targeted training, protocol revision, and model oversight.

Table 2 summarizes this unified functional taxonomy and links each QC layer to its inputs, typical outputs, and clinical value.

Table 1: Direct evidence profile and principal gaps across imaging modalities

Modality	Representative QC tasks	Representative directly relevant study designs	Principal evidence gap
Digital radiography	Projection recognition, anatomical coverage, rotation, inspiration, and foreign objects	Retrospective technical development and internal validation, with limited task-specific external testing	Prospective effects on repeats, diagnostic adequacy, and difficult patient subgroups remain unclear
CT	Automated positioning, scan range, dose, respiratory phase, and reconstruction quality	Dual-center or multivendor retrospective validation and simulated or observational dose analyses	Technical agreement and simulated dose gains are insufficiently linked to diagnostic and patient outcomes
MRI	Motion and artifact detection, research-data screening, cardiac planning, and reacquisition	Cross-dataset or cross-scanner retrospective validation and limited workflow-feasibility studies	Sequence-specific thresholds, external clinical validation, and net benefit of reacquisition remain uncertain
Ultrasound	Standard planes, required anatomy, examination completeness, and multidomain echocardiographic quality	Multicenter real-world implementation and pre–post feedback comparisons	Observational designs cannot separate AI from training and temporal effects; reference standards remain subjective
Mammography	Positioning landmarks, PGMI grading, real-time feedback, and phantom scoring	Longitudinal implementation, multicenter human–AI agreement, and phantom-QC validation	Controlled causal evidence is lacking; holistic grading shows limited agreement with expert consensus
PET/SPECT	Perceptual image quality, uniformity images, and PET phantom analysis	Small patient-image internal validation and larger equipment- or phantom-image series	Patient-outcome validation across tracers, scanners, and reconstruction protocols remains limited

Note. This table summarizes representative directly relevant study designs identified in this targeted narrative review; it is not a formal certainty-of-evidence rating.

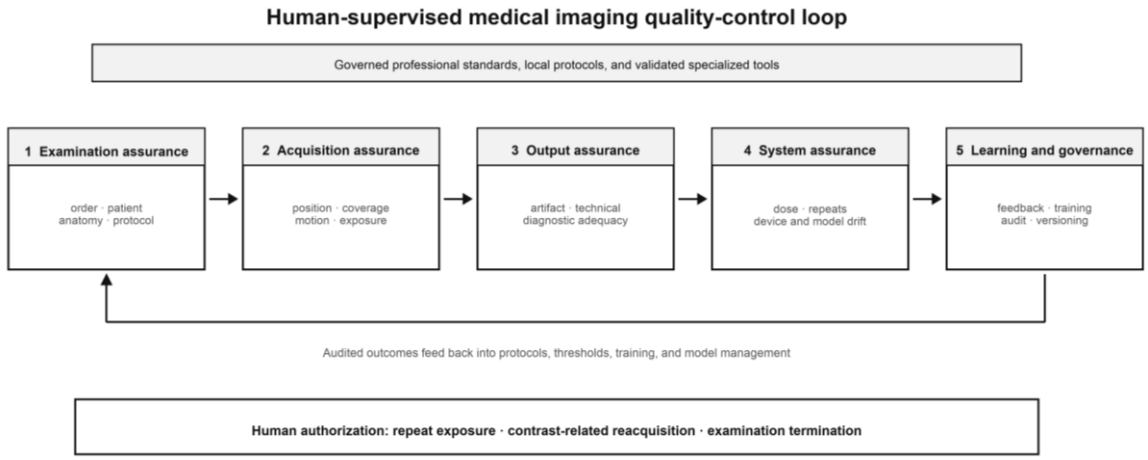

Figure 1: Closed-loop functional framework for human-supervised medical imaging quality control

Table 2: Unified functional taxonomy of AI-enabled medical imaging QC

QC layer	Principal inputs	Representative tasks	Immediate output	Clinical value
Examination assurance	Order, worklist, DICOM metadata, and localizer	Patient–study consistency, projection or series recognition, protocol and completeness checks	Missing or incorrect component alert	Prevent wrong, incomplete, or misrouted examinations
Acquisition assurance	Positioning camera, live image or video, geometry, and exposure data	Centering, positioning, coverage, standard plane, motion, breath hold, and exposure guidance	Real-time corrective feedback	Correct errors before the patient leaves or exposure continues
Output assurance	Projection, series, reconstructed image, or cine loop	Artifact detection, technical adequacy, noise, sharpness, contrast, and task-specific sufficiency	Accept, review, or reacquire recommendation with reason	Improve consistency and reduce delayed recalls
System assurance	Dose reports, repeat logs, phantom images, and device records	Dose outliers, protocol drift, reconstruction surveillance, and detector or phantom QC	Control-chart alert and investigation trigger	Protect equipment performance and longitudinal stability
Learning and governance	QC results, operator response, overrides, and clinical outcomes	Dashboards, targeted education, threshold review, and model-drift monitoring	Individual and departmental improvement plan	Convert isolated errors into sustained improvement

An overall quality score is most clinically useful when it is linked to a specific cause, confidence level, and actionable response. The same low score may imply repositioning, another breath hold, protocol adjustment, equipment investigation, or no reacquisition because the existing images still answer the clinical question. Systems should therefore provide the defect, localization evidence, uncertainty, and recommended action, while preserving a professional route for review and override. In this review, *repeat exposure* refers primarily to another radiographic exposure, *reacquisition* to repetition of a CT, MRI, or ultrasound series or acquisition, and *recall* to bringing a patient back after departure; these actions differ in risk, cost, and evaluation timing. Repeat exposure, contrast-related reacquisition, and examination termination are high-risk actions and should not be executed autonomously.

4. Evidence Across Imaging Modalities

4.1 Digital Radiography

Projection standards and anatomical landmarks are relatively explicit in radiography. Missing anatomy, rotation, inadequate inspiration, scapular overlap, and foreign objects can often be recognized directly. Research has progressed from whole-image adequacy classification toward segmentation, landmark measurement, and defect localization in chest, knee, and elbow imaging [6–11]. These interpretable measurements are more readily converted into positioning feedback than a single score.

Most studies nevertheless remain retrospective development or internal-validation studies, and high image- or projection-level accuracy does not establish adequacy of the complete examination. Data should be partitioned by patient and examination. Prospective evaluation should address unnecessary repeats, missed deficiencies, and diagnostic impact in older adults, bedside imaging, postoperative anatomy, implants, and patients with limited mobility.

4.2 Computed Tomography

CT positioning and scan range affect noise distribution, anatomical coverage, and radiation exposure. Three-dimensional camera or localizer-based methods can reduce off-centering, and automatic scan-range delimitation can achieve high agreement with expert annotations [12–14]. In one coronary CT angiography study, automated scan-range optimization was estimated to reduce effective dose by approximately 10%–13% [15]. Post-acquisition models have also assessed dose outliers, noise and sharpness, respiratory phase, and low-dose image quality [16–19].

Simulated dose reduction, anatomical overlap, and improved numerical image quality do not by themselves demonstrate better diagnostic performance or patient net benefit. Thresholds must be tied to the clinical task and revalidated after changes in scanner, protocol, automatic exposure control, or reconstruction software.

4.3 Magnetic Resonance Imaging

MRI quality depends jointly on sequence, anatomy, coil, field strength, acceleration, patient motion, and reconstruction. Current applications include research-data screening, diffusion-artifact detection, brain-MRI motion assessment, and guidance for replanning or reacquiring cardiac cine sequences [20–25]. Evidence favors artifact- and sequence-specific explanations over a universal quality score, while declining cross-site performance indicates that a single threshold is unlikely to be transportable.

Clinical reacquisition also requires consideration of examination duration, patient tolerance, contrast timing, and whether a defect remains correctable. As AI-based reconstruction becomes common, an independent QC layer is needed to monitor changes in AI-generated images and reconstruction drift [25].

4.4 Ultrasound

Ultrasound is highly operator dependent: image quality cannot be separated from probe manipulation, plane selection, depth, gain, focus, and temporal completeness. This dependence makes ultrasound particularly suitable for immediate guidance. Fetal-ultrasound systems can jointly evaluate standard planes, required anatomy, and examination completeness, whereas echocardiographic systems can assess view, contour, alignment, depth, gain, and cardiac-cycle integrity [26–30].

A first-trimester auditing study reported 7.7% and 5.1% increases in standard images acquired by junior and mid-level operators after feedback [26]. A multicenter real-world study from western China observed improved examination- and plane-level standardization [27]. These findings support

workflow feasibility, but the nonrandomized designs cannot attribute all improvement to AI rather than training, temporal trends, or workforce changes. Because experts may also disagree about acceptable planes, studies should report multi-rater agreement and label uncertainty.

4.5 Mammography

Mammographic positioning criteria are relatively standardized. Models can assess nipple depiction, posterior nipple line, pectoral muscle, inframammary fold, tissue cutoff, skin folds, and compression [31,32]. A single-center longitudinal study reported a decline in inadequate examinations from 13.31% to 3.20% at a later time point after deployment of real-time AI feedback [33]. This finding associates continuous feedback with improved practice, but the before–after design and relationships between some authors and the software provider require cautious causal interpretation.

A multicenter comparison found only slight-to-fair agreement between overall AI-derived PGMI grading and expert consensus [34]. This negative result is important: objective landmarks may support assessment without reproducing a holistic quality category. Automated phantom-image scoring may improve equipment-QC efficiency, but metrological traceability, software versioning, and periodic human verification remain necessary [35].

4.6 PET, SPECT, and Nuclear Medicine

Nuclear medicine QC includes patient preparation, injected activity, uptake time, motion, attenuation correction, reconstruction, quantitation, and detector performance. A small PET study supports the feasibility of automated perceptual-quality grading [36]. Automated analysis of gamma-camera uniformity images and ACR PET phantom images can improve anomaly detection and quantitative consistency [37,38].

External and prospective patient-image evidence remains more limited than in radiography or ultrasound. Future studies should link automated scores to standardized physics QC, quantitative-biomarker stability, and diagnostic performance across tracers, scanners, and reconstruction protocols.

5. Evaluation Requirements for AI-QC Systems

5.1 Intended Use and Unit of Analysis

The intended use should specify whether a system supports immediate reacquisition, retrospective audit, staff education, research-data screening, or equipment surveillance. The unit may be a frame, image, projection, series, complete examination, device-day, or operator-month. Performance estimated at image level may overstate effectiveness when action is taken at patient or examination level.

5.2 Reference Standard and Label Reliability

Studies should report the quality rubric, rater experience and training, blinding, number of raters, consensus or adjudication, and inter- and intrarater agreement. Labels should distinguish technical imperfection from inability to answer the clinical question. When no uncontested gold standard exists, investigators should state whether labels arose from one expert, consensus, adjudication, or a probabilistic distribution of multiple ratings and should examine sensitivity to alternative reference-standard definitions. For ordinal or continuous ratings, weighted kappa, intraclass correlation, or Bland–Altman agreement is more informative than correlation alone.

5.3 Data Independence and Statistical Analysis

Random image-level splitting causes leakage when multiple images, series, or examinations belong to one patient. Development and testing should be separated at least by patient, with temporal, institutional, device, and protocol validation where possible. Analyses should account for images nested within examinations, examinations within patients, and patients within operators or institutions by using multilevel models, cluster-robust errors, or stratified bootstrap procedures.

5.4 Metrics Linked to Clinical Action

Continuous scores require agreement and calibration. Deployment studies should measure latency, alert rate, alerts per effective correction, overrides, QC time, repeat or recall rates, dose, diagnostic confidence or performance, and downstream failures. A system intended to replace or reduce human review requires a prespecified clinically acceptable noninferiority margin. Systems that trigger action may also require decision-curve or net-benefit analysis so that statistical significance is not mistaken for clinical utility. Table 3 presents a minimum evaluation set.

5.5 Stage-gated Clinical Translation

Retrospective testing establishes technical potential only. A locked model should first undergo external and temporal validation, followed by local silent operation to assess routing, latency, calibration, and failure handling. Only then should it enter a limited, reversible supervised pilot. Comparative evaluation should address clinical effectiveness, after which post-deployment monitoring should cover data and performance drift, subgroup differences, user response, and actual quality outcomes [2,4,39–41].

Table 3: Minimum evaluation set for a clinically credible AI-QC system

Domain	Core question	Minimum recommended evidence
Intended use	Which decision, user, timing, and risk does the output support?	Predefined use case, unit of analysis, contraindications, and failure handling
Reference standard	Is quality reproducible and clinically meaningful?	Explicit rubric, rater training and blinding, consensus or adjudication, and inter- and intrarater agreement
Data independence	Are test cases independent of development data?	Patient-level separation, temporal and external testing, and transparent handling of exclusions, missing data, and repeat studies

Domain	Core question	Minimum recommended evidence
Statistical design	Are repeated images and hierarchical clustering handled?	Performance at the decision unit with multilevel models, cluster-robust errors, or stratified bootstrap confidence intervals
Performance and calibration	Does the system identify the target defect at the operating threshold?	Prevalence, sensitivity, specificity, predictive values, calibration, error analysis, and confidence intervals
Robustness and subgroups	Is performance stable across sites, devices, protocols, and patients?	External validation, prespecified subgroup analyses, stress testing, and out-of-distribution handling
Workflow	Does the system change action without excessive burden?	Latency, alert rate, overrides, QC time, repeat or recall rates, user studies, and unnecessary actions
Clinical value	Does technical improvement preserve or improve diagnostic utility and safety?	Task-based image quality, diagnostic confidence or performance, dose, examination time, net benefit, prespecified noninferiority margins, and patient-relevant outcomes
Life cycle	Can failure, drift, and new human-factor risks be detected after deployment?	Versioning, audit logs, integrity checks, data and performance monitoring, incident reporting, and suspension or retirement procedures

Note. AI, artificial intelligence; QC, quality control.

6. From Isolated Models to a Human-Supervised QC Agent

This review defines a QC agent as a conceptual and testable governed software layer that invokes multiple locked and independently validated specialized tools; it is not a large model that replaces professional standards. Inputs may include the order, patient and protocol context, DICOM metadata, dose reports, images or video, and device identifiers. The tool layer may provide view recognition, segmentation, landmark measurement, artifact detection, perceptual-quality assessment, dose rules, and equipment control charts.

Decision mapping should use version-controlled professional standards, local protocols, and preapproved thresholds. The interaction layer should communicate location, evidence, recommended action, and uncertainty. The audit layer should retain input provenance, model version, output, user response, override reason, repeat outcome, and longitudinal performance. Language or vision–language models may translate structured results into understandable feedback, but they should not invent QC criteria, override deterministic safety rules, or issue high-risk reacquisition commands independently.

Low confidence, tool disagreement, out-of-distribution inputs, and device or protocol changes should interrupt automated recommendation and escalate the case for human review. The stage-gated pathway in Figure 2 is a design proposal. Prospective comparative studies are required to determine whether this orchestrated architecture provides net benefit over isolated tools or conventional manual QC.

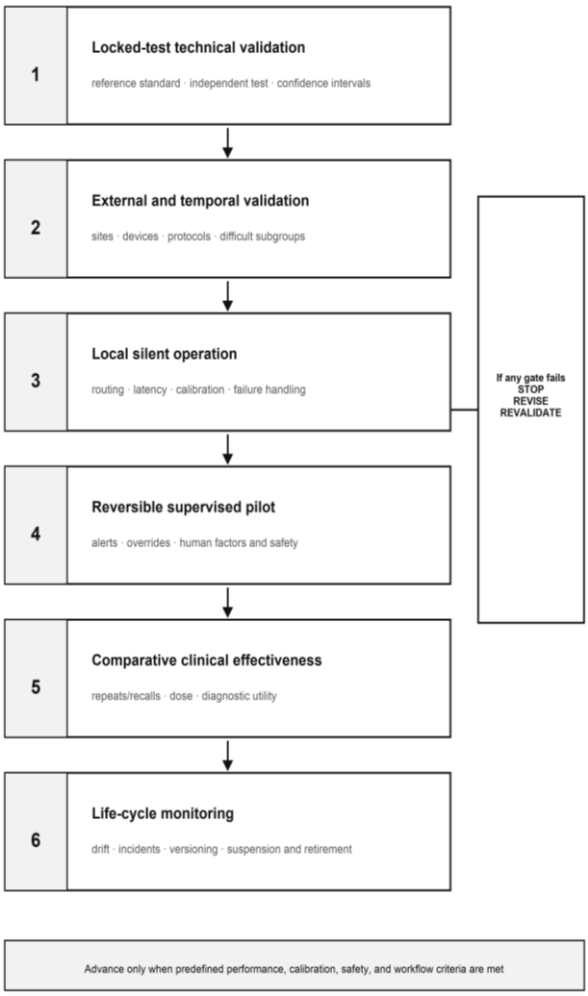

Figure 2: Stage-gated clinical translation pathway for an AI quality-control system.

7. Barriers and Research Priorities

First, quality labels often mix technical appearance, diagnostic utility, and personal preference. Multidisciplinary, multicenter work should develop modality-specific ontologies and rubrics that link each defect to a clinical consequence and corrective action.

Second, vendor, post-processing, protocol, population, and operator differences create domain shift. Claims of clinical readiness should require at least temporal or external validation and prespecified analysis of difficult subgroups, including older adults, children, large body habitus, limited mobility, bedside imaging, implants, and postoperative anatomy.

Third, improved AI scores do not necessarily reduce missed diagnoses, unnecessary repeats, or patient burden. Studies should compare human-only, AI-only, and human–AI workflows using cluster randomized, stepped-wedge, controlled before–after, or interrupted time-series designs appropriate for workflow interventions.

Fourth, automation bias and alert burden can introduce new errors. Interfaces should show cause, confidence, and alternative actions. Institutions should sample concordant and discordant cases and monitor overrides, alert fatigue, and unnecessary reacquisition.

Fifth, models are affected by scanner replacement, software upgrades, protocol revision, and population change. Institutions need named ownership, version control, data- and performance-drift monitoring, incident reporting, and predefined triggers for recalibration, suspension, or retirement.

8. Limitations of This Review

This targeted structured narrative review searched only PubMed and may have missed studies indexed primarily in IEEE Xplore, Embase, or engineering conference proceedings. The 2017 start date focuses on contemporary deep learning but may underrepresent earlier automated equipment QC. The review did not use duplicate screening, formal risk-of-bias assessment, or quantitative synthesis; Table 1 is therefore an evidence profile rather than a certainty grade. The scope was deliberately limited to six diagnostic imaging categories, and the conclusions should not be extrapolated to pathology, ophthalmology, radiotherapy, or interventional imaging. Rapidly emerging publications from 2025–2026 include online-ahead-of-print and observational implementation studies whose durability, reproducibility, and bias remain to be established.

9. Conclusion

AI-enabled imaging QC is progressing from retrospective single-defect classification toward explainable, cause-specific assessment, feedback during acquisition, and post-deployment surveillance. Clinical value is most plausible when the criterion is explicit, the error remains correctable, and the outcome is measured at workflow level. However, evidence for clinical effectiveness in radiography, CT, and MRI remains limited, and real-world improvements in ultrasound and mammography are still based mainly on observational studies.

Future systems should not pursue one universal quality score. They should coordinate validated specialized tools under human supervision, interpret results against governed standards, and preserve audit and exit mechanisms. This agent architecture is plausible, but its incremental benefit over isolated models and conventional manual QC still requires multicenter, prospective, workflow-centered evaluation.

Fund Project

Platform: Xianyang Polytechnic Institute (Medical School), Number: 2026ZCTD06, Research and Application of an AI-Agent-Based Quality Control System for X-ray Imaging Examinations (Start and End Time: June 2026 - June 2027). Responsible Person: He Zeming, Participants: Qiu Yue, Guo Qiuting, Zhang Mi, Xu Xuejie. Project Support: Traditional Chinese Medicine Health and Wellness Innovation Team for Chronic Diseases in the Elderly.

References

[1] Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. Radiol Artif Intell. 2024;6(4):e240300. doi: 10.1148/ryai.240300.

[2] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022; 377: e070904. doi: 10.1136/bmj-2022-070904.

[3] Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31(10):3283-3289. doi: 10.1038/s41591-025-03953-8.

[4] Brady AP, Allen B, Chong J, Kotter E, Kottler N, Mongan J, et al. Developing, Purchasing, Implementing and Monitoring AI Tools in Radiology: Practical Considerations. A Multi-Society Statement from the ACR, CAR, ESR, RANZCR and RSNA. Radiol Artif Intell. 2024;6(1):e230513. doi: 10.1148/ryai.230513.

[5] Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019; 4: 5. doi: 10.1186/s41073-019-0064-8.

[6] Nousiainen K, Mäkelä T, Piilonen A, Peltonen JI. Automating chest radiograph imaging quality control. Phys Med. 2021; 83: 138-145. doi: 10.1016/j.ejmp.2021.03.014.

[7] Hu J, Zhang C, Zhou K, Gao S. Chest X-Ray Diagnostic Quality Assessment: How Much Is Pixel-Wise Supervision Needed? IEEE Trans Med Imaging. 2022; 41(7): 1711-1723. doi: 10.1109/TMI.2022.3149171.

[8] Wang Q, Han X, Song L, Zhang X, Zhang B, Gu Z, et al. Automatic quality assessment of knee radiographs using knowledge graphs and convolutional neural networks. Med Phys. 2024;51(10):7464-7478. doi: 10.1002/mp.17316.

[9] Lai Q, Chen W, Ding X, Huang X, Jiang W, Zhang L, et al. Quality control of elbow joint radiography using a YOLOv8-based artificial intelligence technology. Eur Radiol Exp. 2024;8(1):107. doi: 10.1186/s41747-024-00504-7.

[10] Ma C, Peng R, Li B, Zhang D, Zhang R, Zhang Z, et al. Application of multi-scale feature extraction and explainable machine learning in chest x-ray position evaluation within an integrated learning framework. Eur Radiol. 2026;36(4):3143-3157. doi: 10.1007/s00330-025-12097-9.

[11] Sorace L, Raju N, O'Shaughnessy J, Kachel S, Jansz K, Yang N, et al. Assessment of inspiration and technical quality in anteroposterior thoracic radiographs using machine learning. Radiography (Lond). 2024; 30(1): 107-115. doi: 10.1016/j.radi.2023.10.014.

[12] Gang Y, Chen X, Li H, Wang H, Li J, Guo Y, et al. A comparison between manual and artificial intelligence-based automatic positioning in CT imaging for COVID-19 patients. Eur Radiol. 2021; 31(8): 6049-6058. doi: 10.1007/s00330-020-07629-4.

[13] Salimi Y, Shiri I, Akavanallaf A, Mansouri Z, Arabi H, Zaidi H. Fully automated accurate patient positioning in computed tomography using anterior-posterior localizer images and a deep neural network: a dual-center study. Eur Radiol. 2023;33(5):3243-3252. doi: 10.1007/s00330-023-09424-3.

[14] Demircioğlu A, Kim MS, Stein MC, Guberina N, Umutlu L, Nassenstein K. Automatic Scan Range Delimitation in Chest CT Using Deep Learning. Radiol Artif Intell. 2021;3(3):e200211. doi: 10.1148/ryai.2021200211.

[15] Demircioğlu A, Bos D, Demircioğlu E, Qaadan S, Glasmachers T, Bruder O, et al. Deep learning-based scan range optimization can reduce radiation exposure in coronary CT angiography. Eur Radiol. 2024;34(1):411-421. doi: 10.1007/s00330-023-09971-9.

[16] Meineke A, Rubbert C, Sawicki LM, Thomas C, Klosterkemper Y, Appel E, et al. Potential of a machine-learning model for dose optimization in CT quality assurance. Eur Radiol. 2019;29(7):3705-3713. doi: 10.1007/s00330-019-6013-6.

[17] Chun M, Choi JH, Kim S, Ahn C, Kim JH. Fully automated image quality evaluation on patient CT: Multi-vendor and multi-reconstruction study. PLoS One. 2022;17(7):e0271724. doi: 10.1371/journal.pone.0271724.

[18] Su J, Li M, Lin Y, Xiong L, Yuan C, Zhou Z, et al. A deep learning-based automatic image quality assessment method for respiratory phase on computed tomography chest images. Quant Imaging Med Surg. 2024; 14(3): 2240-2254. doi: 10.21037/qims-23-1273.

[19] Lee W, Wagner F, Galdran A, Shi Y, Xia W, Wang G, et al. Low-dose computed tomography perceptual image quality assessment. Med Image Anal. 2025;99:103343. doi: 10.1016/j.media.2024.103343.

[20] Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One. 2017;12(9):e0184661. doi: 10.1371/journal.pone.0184661.

[21] Samani ZR, Alappatt JA, Parker D, Ismail AAO, Verma R. QC-Automator: Deep Learning-Based Automated Quality Control for Diffusion MR Images. Front Neurosci. 2019;13:1456. doi: 10.3389/fnins.2019.01456.

[22] Lei K, Syed AB, Zhu X, Pauly JM, Vasanawala SS. Artifact- and content-specific quality assessment for MRI with image rulers. Med Image Anal. 2022; 77: 102344. doi: 10.1016/j.media.2021.102344.

[23] Cheung HC, Vimalesvaran K, Zaman S, Michaelides M, Shun-Shin MJ, Francis DP, et al. Automating quality control in cardiac magnetic resonance: Artificial intelligence for discriminative assessment of planning and motion artifacts and real-time reacquisition guidance. J Cardiovasc Magn Reson. 2024; 26(2): 101067. doi: 10.1016/j.jocmr.2024.101067.

[24] Manso Jimeno M, Ravi KS, Fung M, Oyekunle D, Ogbole G, Vaughan JT, et al. Automated detection of motion artifacts in brain MR images using deep learning. NMR Biomed. 2025;38(1):e5276. doi: 10.1002/nbm.5276.

[25] White OA, Shur J, Castagnoli F, Charles-Edwards G, Whitcher B, Collins DJ, et al. Quantitative image quality metrics enable resource-efficient quality control of clinically applied AI-based reconstructions in MRI. MAGMA. 2025;38(3):547-560. doi: 10.1007/s10334-025-01253-3.

[26] Cao X, Li B, Zhou Y, Cao Y, Yang X, Hu X, et al. Effectiveness and clinical impact of using deep learning for first-trimester fetal ultrasound image quality auditing. BMC Pregnancy Childbirth. 2025;25(1):375. doi: 10.1186/s12884-025-07485-4.

[27] Zhao J, Tang Y, Li S, Wang K, Tao J, Chen C, et al. AI-based quality control was associated with improved fetal ultrasound image quality in low-resource settings: a real-world multicenter study from West China. BMC Med. 2026;24(1):235. doi: 10.1186/s12916-026-04768-1.

[28] Li X, Liao L, Wu K, Meng AT, Jiang Y, Zhu Y, et al. An automatic and real-time echocardiography quality scoring system based on deep learning to improve reproducible assessment of left ventricular ejection fraction. Quant Imaging Med Surg. 2025;15(1):770-785. doi: 10.21037/qims-24-512.

[29] Zhang J, Feng B, Cheng Y, Zhou F, Lu X, Zhu X, et al. Artificial intelligence-assisted quality assessment of mid-trimester ultrasound examinations using large vision-language models. BMC Pregnancy Childbirth. 2026;26(1):614. doi: 10.1186/s12884-026-09073-6.

[30] Shi Z, Cheng H, Qi Z, Shan C, Taub CC, Ouyang D, et al. A Clinically Interpretable AI System for Real-Time Quality Control of Transthoracic Echocardiography: Development, Validation, and Deployment. J Am Soc Echocardiogr. 2026:S0894-7317(26)00307-X. doi: 10.1016/j.echo.2026.06.007.

[31] Brahim M, Westerkamp K, Hempel L, Lehmann R, Hempel D, Philipp P. Automated Assessment of Breast Positioning Quality in Screening Mammography. Cancers (Basel). 2022;14(19):4704. doi: 10.3390/cancers14194704.

[32] Eby PR, Martis LM, Paluch JT, Pak JJ, Chan AHL. Impact of Artificial Intelligence-driven Quality Improvement Software on Mammography Technical Repeat and Recall Rates. Radiol Artif Intell. 2023; 5(6):e230038. doi: 10.1148/ryai.230038.

[33] Sexauer R, Riehle F, Borkowski K, Ruppert C, Potthast S, Schmidt N. Enhancing breast positioning quality through real-time AI feedback. Eur Radiol. 2026; 36(1): 55-63. doi: 10.1007/s00330-025-11812-w.
[34] Santner T, Tardy M, Stalheim JG, Frei S, Santner W, Gianolini S, et al. AI-based image quality assessment of positioning in mammography: considerations and challenges. Insights Imaging. 2026;17(1):47. doi: 10.1186/s13244-025-02191-3.
[35] Yun H, Noh S, Cho H, Ko EY, Yang Z, Woo OH. AI-Driven quality assurance in mammography: Enhancing quality control efficiency through automated phantom image evaluation in South Korea. PLoS One. 2025;20(9):e0330091. doi: 10.1371/journal.pone.0330091.
[36] Zhang H, Liu Y, Wang Y, Ma Y, Niu N, Jing H, et al. Deep learning model for automatic image quality assessment in PET. BMC Med Imaging. 2023;23(1):75. doi: 10.1186/s12880-023-01017-2.
[37] Nelson JS, Samei E. Automated quality control in nuclear medicine using the structured noise index. J Appl Clin Med Phys. 2020;21(4):80-86. doi: 10.1002/acm2.12850.
[38] DiFilippo FP, Patel M, Patel S. Automated Quantitative Analysis of American College of Radiology PET Phantom Images. J Nucl Med Technol. 2019; 47(3): 249-254. doi: 10.2967/jnmt.118.221317.
[39] Kore A, Abbasi Bavil E, Subasri V, Abdalla M, Fine B, Dolatabadi E, et al. Empirical data drift detection experiments on real-world medical imaging data. Nat Commun. 2024;15(1):1887. doi: 10.1038/s41467-024-46142-w.
[40] Ramwala OA, Lowry KP, Cross NM, Hsu W, Austin CC, Mooney SD, et al. Establishing a Validation Infrastructure for Imaging-Based Artificial Intelligence Algorithms Before Clinical Implementation. J Am Coll Radiol. 2024;21(10):1569-1574. doi: 10.1016/j.jacr.2024.04.027.
[41] Eche T, Schwartz LH, Mokrane FZ, Dercle L. Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: Role of Computation Stress Testing to Overcome Underspecification. Radiol Artif Intell. 2021;3(6):e210097. doi: 10.1148/ryai.2021210097.

Appendix 1. PubMed Search Strategy

Last search date: August 4, 2026. Search string: (("artificial intelligence" [Title/Abstract] OR "machine learning" [Title/Abstract] OR "deep learning" [Title/Abstract]) AND ("quality control" [Title/Abstract] OR "quality assurance" [Title/Abstract] OR "image quality" [Title/Abstract] OR positioning [Title/Abstract] OR "standard plane" [Title/Abstract] OR artifact [Title/Abstract] OR artifacts [Title/Abstract] OR "scan range" [Title/Abstract]) AND (radiograph [Title/Abstract] OR radiographs [Title/Abstract] OR radiography [Title/Abstract] OR "x-ray" [Title/Abstract] OR "computed tomography" [Title/Abstract] OR "magnetic resonance" [Title/Abstract] OR ultrasound [Title/Abstract] OR mammogram [Title/Abstract] OR mammograms [Title/Abstract] OR mammography [Title/Abstract] OR PET [Title/Abstract] OR SPECT [Title/Abstract] OR "nuclear medicine" [Title/Abstract])) AND ("2017/01/01" [Date - Publication]: "2026/08/04"[Date - Publication]) AND English [Language]. This broad query was used to identify candidate literature and was supplemented by modality-specific combinations and backward citation tracing. Because this was a targeted narrative review, it is not presented as an exhaustive PRISMA-style duplicate-screening process.
[42]