www.ijcbhsr.com/

INTERNATIONAL JOURNAL OF CLINICAL, BIOMEDICAL AND HEALTH SCIENCES RESEARCH
2026, Vol 1 Issue 1, pp. 16-19
ISSN:xxxx-xxxx
Open Access. Online

Artificial Intelligence in Clinical Diagnosis

Dr. Alexander James Moore

Associate Professor, Department of Business Finance, Royal Metropolitan University, Edinburgh, United Kingdom.

International Journal of Clinical, Biomedical and Health Sciences Research

vol 1, Issue1, (2026).

Received: Jan 14, 2026.
Accepted: April 24, 2026.
Published: June 4, 2026.

ABSTRACT:
Artificial intelligence (AI) can identify patterns in medical images, physiological signals and clinical records, offering support for screening, triage and selected diagnostic decisions. Its usefulness depends on the clinical question, patient population, reference standard, workflow and consequences of false results. This narrative review discusses diagnostic applications, study design, performance measures, external validation and ethical oversight. Published results from a prospective trial of an autonomous diabetic retinopathy system illustrate sensitivity, specificity and imageability in a defined primary-care setting. Those metrics cannot be generalized to other diseases or devices. Evidence from AI-assisted mammography shows how a system can be studied in a randomized screening workflow, while regulatory authorization remains tied to a specific intended use. Safe deployment requires data-quality checks, clear clinician and patient communication, referral pathways, monitoring for drift and evaluation of outcomes across patient groups.

Keywords: artificial intelligence; clinical diagnosis; diagnostic accuracy; diabetic retinopathy; mammography; external validation; patient safety

1 Introduction

Clinical diagnosis combines history, examination, tests and professional judgment to estimate what condition is present and what action is appropriate. AI systems may classify an image, flag an abnormality, prioritize a worklist or summarize information. These are distinct uses. A model that detects a retinal feature is not automatically an autonomous diagnosis, and a high test-set accuracy does not prove improved patient outcomes. The World Health Organization (WHO) calls for human rights, accountability and public benefit in the development and use of AI for health [1].

This article is a narrative review; it presents no newly collected patient data. Its central question is how to interpret a reported diagnostic result before adopting it in a particular care pathway. The answer depends on representative validation, action thresholds, follow-up capacity and ongoing surveillance after deployment.

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

2 Diagnostic applications and workflow

Medical imaging has supplied many well-defined AI tasks, including detection of diabetic retinopathy in fundus images and assistance in mammography screening. Other applications include electrocardiogram interpretation, pathology image analysis and warning systems using electronic health records. Input type, disease prevalence and clinical setting differ across these tasks; a single aggregate performance claim therefore has little meaning without a target population and reference standard.



Figure 1. A diagnostic AI workflow with data-quality control, defined review and follow-up.

A responsible pathway begins with eligibility. Some patients have an image or signal that the system cannot assess, and an uninterpretable result needs a safe next step. Input quality checks should flag poor focus, missing views, incompatible devices or out-of-range values. The output should state the intended decision, its limitations and any uncertainty. A referral recommendation only helps patients when appointment access, communication and tracking are built into the service.

Roles must be explicit. A clinician-assistive tool asks a trained professional to consider a result; an authorized autonomous system may produce a result for a defined use under specified conditions. Clinicians still need to manage symptoms, conflicting evidence and cases beyond the device indication. Regulators evaluate particular devices and intended uses. The US Food and Drug Administration (FDA) maintains a list of AI-enabled devices authorized for marketing, which should not be read as approval of every possible use of those devices [2].

3 Performance measures

Sensitivity is the proportion of people with the target condition who test positive; specificity is the proportion without it who test negative. Positive predictive value depends on prevalence, so it may change markedly when the same model moves from a specialty clinic to screening. Calibration matters when a system reports probabilities. Decision thresholds should reflect the harms of missed disease, unnecessary referral and delay. Imageability or test completion must also be counted: excluding unusable cases can make reported accuracy appear better than the real pathway experienced by patients.

4 Evidence from selected clinical studies

Table 1. Selected evidence and its limits. The studies address different questions and should not be pooled.

Application	Study design and setting	Key interpretation
Diabetic retinopathy	Prospective pivotal trial of an autonomous system in primary-care offices [3]	Sensitivity 87.2%, specificity 90.7%, imageability 96.1% for a defined target and workflow.
Mammography	Randomized MASAI screening study of AI-supported reading versus	Screen-detected cancers and reading workload evaluated in a specific program.

	standard double reading [4]	
Other image or signal tasks	Task-specific development and external cohorts	Validate device, population, reference standard and downstream clinical action separately.

Abràmoff and colleagues reported a prospective pivotal trial of an autonomous system for detecting more-than-mild diabetic retinopathy in primary-care offices. It achieved 87.2% sensitivity, 90.7% specificity and 96.1% imageability against the study reference standard [3]. The graph presents these three published metrics. It does not compare the AI with every ophthalmologist, and its values should not be transplanted to other populations, cameras or retinal diseases without evidence.

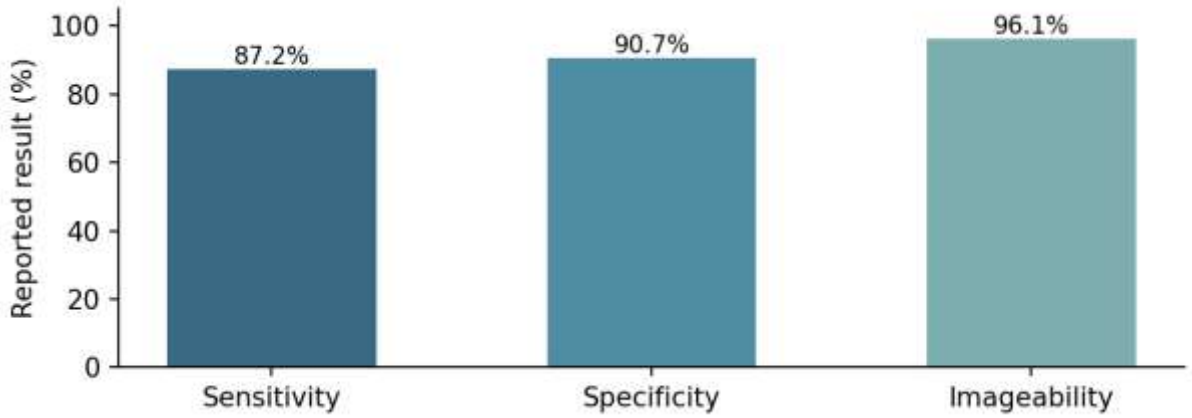


Figure 2. Three reported endpoints from the diabetic retinopathy pivotal trial [3]. These are published study results, not new measurements.

The MASAI trial randomized screening participants to AI-supported versus standard double reading. Its clinical safety analysis reported 261 versus 250 screen-detected cancers in the respective groups, with interpretation tied to its screening design and predefined comparisons [4]. Detection count alone does not tell the whole story: interval cancers, false positives, recall rates and workload need assessment over appropriate follow-up. A favorable workflow result in one screening program does not establish the same balance in a different population.

5 Validation, bias and uncertainty

Development data often differ from real clinical practice. Images may come from different scanners, hospitals or disease spectra. Labels can reflect expert disagreement or inconsistent follow-up. External validation should use independent sites and time periods, report missing or unusable data and examine performance by age, sex, ethnicity and relevant clinical subgroups where feasible. Prospective evaluation then tests whether the tool changes real decisions and outcomes rather than merely reproducing retrospective labels.

Data leakage is a common risk: a model may learn artifacts linked to the diagnosis or receive information that would not be available at decision time. A random split at the image level can place images from the same patient in both training and testing, inflating results. Site-level and temporal splits, locked models, prespecified thresholds and transparent reporting reduce this problem. Confidence intervals convey sampling uncertainty but do not capture every deployment shift.

False negatives may delay treatment; false positives may cause anxiety, procedures and unnecessary visits. Automation bias can lead staff to over-trust a plausible output, while alert fatigue can cause useful signals to be ignored. A clinician should know when to override an AI result and how to document

disagreement. Training must explain intended use and failure modes in plain language, not only interface operation.

6 Governance and patient communication

The WHO framework emphasizes autonomy, safety, transparency, responsibility, inclusiveness and sustainability [1]. These principles translate into practical requirements: governance of training data and access, clear accountability for decisions, privacy protection and monitoring for unequal performance. Patients should receive an understandable explanation of the AI system's role, particularly when it affects referral or care. Local legal requirements for consent and medical devices vary, so implementation requires jurisdiction-specific review. Post-deployment monitoring should track input quality, result distributions, referral completion, false results identified later, clinician overrides and patient complaints. Model drift may follow equipment replacement, changing disease prevalence or altered workflows. Recalibration or model updates require a defined change-control and revalidation process. A service must preserve a fallback route during device outage or poor input quality.

7 Future research priorities

Comparative studies should examine patient-important outcomes, not only area under a curve. Useful endpoints include time to diagnosis, avoidable vision loss or other disease-specific harm, equitable access, referral completion and resource use. Multi-site trials should include community and resource-constrained settings. Interoperable records and careful prospective follow-up can reveal whether an apparent increase in detection improves treatment or simply shifts referrals to overwhelmed specialists.

Foundation models may perform several tasks, but broad capability raises new validation questions. A model that generates fluent differential diagnoses can still omit a critical condition or cite nonexistent evidence. Researchers should test each intended function, separate generation from verified clinical facts and audit updates. Procurement should ask for independent evaluation in the local patient population and a credible plan for monitoring.

8 Conclusion

AI can improve clinical diagnosis when a narrowly defined tool is tested in the setting where it will be used and embedded in a safe care pathway. Published retinal and mammography studies show why prospective evidence, usable inputs and meaningful clinical endpoints matter. Performance figures alone cannot substitute for external validation, clinician oversight, patient communication and follow-up. Responsible adoption measures whether patients receive more timely and equitable care, then continues to check safety as populations and workflows change.

References

[1] World Health Organization. (2021). Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200

[2] US Food and Drug Administration. List of Artificial Intelligence-Enabled Medical Devices. https://www.fda.gov/medical-devices/artificial-intelligence-enabled-medical-devices/list-artificial-intelligence-enabled-medical-devices

[3] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine, 1, 39. https://doi.org/10.1038/s41746-018-0040-6

[4] Lång, K., et al. (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): A clinical safety analysis. The Lancet Oncology. https://doi.org/10.1016/S1470-2045(23)00298-X