Journal of Advanced Medical and Dental Sciences Research

@Society of Scientific Research and Studies *NLM ID: 101716117*

Journal home page: www.jamdsr.com doi: 10.21276/jamdsr Indian Citation Index (ICI) Index Copernicus value = 100

(e) ISSN Online: 2321-9599; (p) ISSN Print: 2348-6805

Original Research

Role Of Artificial Intelligence In Medical Diagnostics: A Database Research

Ifath Nazia Ghori

Assistant Professor, Department of Computer Science, Jazan University , Saudi Arabia

ABSTRACT:

Background: Artificial intelligence (AI) has rapidly emerged as a transformative tool in medical diagnostics, driven by increasing availability of digital health data, advances in machine learning algorithms, and demand for efficient diagnostic decision support. AI-based systems are now being applied across multiple diagnostic domains, particularly in imaging, pathology, and electronic health record–driven risk prediction.

Objective: To systematically synthesize evidence from major biomedical databases on the role, diagnostic performance, and clinical implementation of AI-based tools in medical diagnostics.

Methods: A structured database research approach was conducted using PubMed and Scopus. Peer-reviewed studies, systematic reviews, meta-analyses, and regulatory reports evaluating AI for diagnostic applications were included. Data were extracted on study design, diagnostic tasks, performance metrics, validation strategies, and implementation considerations. Findings were synthesized descriptively and organized into thematic evidence tables.

Results: Evidence indicates that AI systems demonstrate high diagnostic accuracy for narrowly defined tasks, particularly in radiology and ophthalmology, where deep learning models achieve performance comparable to specialist interpretation in controlled validation studies. Meta-analyses report high sensitivity for AI-assisted screening in diabetic retinopathy and breast cancer detection. However, substantial heterogeneity exists in study design, dataset representativeness, and external validation. Implementation barriers include dataset shift, algorithmic bias, limited transparency, and incomplete integration into clinical workflows. Regulatory approvals for AI-enabled diagnostic devices have increased steadily, yet post-market surveillance data on real-world effectiveness remain limited.

Conclusion: AI has significant potential to enhance medical diagnostics by improving detection accuracy, standardizing interpretation, and optimizing clinical workflow. Nevertheless, successful clinical adoption requires rigorous external validation, transparent reporting, continuous performance monitoring, and alignment with emerging regulatory frameworks. Future research should prioritize multicenter prospective evaluations and patient-centered outcome assessment.

Keywords: Artificial intelligence; Medical diagnostics; Machine learning; Deep learning; Diagnostic accuracy; Radiology; Ophthalmology; Clinical decision support; Validation; Regulatory governance.

Received: 08 December, 2025 Accepted: 10 January, 2026 Published: 15 January, 2026

Corresponding Author: Ifath Nazia Ghori, Assistant Professor, Department of Computer Science, Jazan University , Saudi Arabia

This article may be cited as: Ghori IN, Role of Artificial Intelligence In Medical Diagnostics: A Database Research. J AdvMed Dent Scie Res 2026; 14(1):41-46.

INTRODUCTION

Medical diagnostics is increasingly data-intensive, driven by high-throughput imaging, expanding laboratory panels, and longitudinal electronic health records (EHRs). This growth has created a dual challenge: clinicians must integrate more information than ever, while health systems face shortages in specialized diagnostic expertise. AI methods—especially DL architectures—have therefore been positioned as "augmentation" tools to improve sensitivity for detection tasks, reduce variability, and accelerate triage, rather than replacing clinicians. In practice, many diagnostic AI systems function as pattern-recognition engines trained to map input data (images, waveforms, or tabular clinical features) to clinically meaningful outputs such as probability of disease, lesion localization, or severity classification.

A defining feature of diagnostic AI is task specificity. Systems built for narrow endpoints (e.g., detection of diabetic retinopathy from fundus photographs, or malignancy suspicion on mammography) often demonstrate strong performance in retrospective test

sets. Landmark clinical-grade AI work in ophthalmology established that DL algorithms could detect diabetic retinopathy at high accuracy using retinal images, catalyzing broader interest in regulated diagnostic AI products. [1] Similarly, large-scale evaluation of AI for breast cancer screening has shown the potential to reduce false negatives and false positives in well-curated screening datasets, suggesting value in population screening workflows. [2] Multiple radiology studies have also reported that AI tools can improve reader performance or workflow efficiency for selected tasks such as mammography interpretation and chest imaging triage. [3]

The COVID-19 pandemic provided an additional "stress test" for diagnostic AI, accelerating both model development and clinical interest. Meta-analytic evidence indicates that chest CT has relatively high diagnostic sensitivity for COVID-19 in certain settings, while later comparative meta-analyses have suggested DL models may achieve high sensitivity with specificity comparable to clinicians for CT-based COVID-19 classification (noting important limitations in study design and data leakage risk). [4,5] These experiences reinforced a core lesson: apparent model performance can inflate when datasets are not representative, when patient selection is biased, or when evaluation lacks robust external validation.

Beyond performance, diagnostic AI adoption depends on regulation, governance, and clinical integration. In the United States, the Food and Drug Administration (FDA) maintains an updated list of AI-enabled medical devices, reflecting the expanding set of cleared/approved tools, especially in radiology. [6] Regulatory science analyses of FDA-authorized ML-enabled devices further indicate rapid year-on-year growth and concentration in specific modalities and pathways. [7] In parallel, the European Union's AI Act entered into force in August 2024 and establishes a risk-based framework with staged applicability timelines, including obligations for high-risk AI systems used in healthcare and regulated products. [8]

Finally, the research community has recognized that inconsistent reporting and incomplete transparency undermine clinical trust. The CONSORT-AI and SPIRIT-AI extensions were developed to improve reporting quality for clinical trials and protocols evaluating AI interventions, aligning evidence generation with patient safety and reproducibility requirements. [9,10] Against this background, a "database research" synthesis is useful to map where evidence is strongest, where it is fragile, and what implementation barriers are most consistently reported.

MATERIALS AND METHODS

Study design

This study was designed as a structured database research synthesis (rapid evidence mapping) focusing on AI applications in **medical diagnostics**. The approach emphasized (i) evidence quality, (ii) diagnostic performance metrics, and (iii) real-world deployment considerations (bias, generalizability, workflow integration, and regulation).

Data sources and search strategy

Two bibliographic databases were selected to balance clinical coverage and multidisciplinary indexing:

- **PubMed/MEDLINE** (biomedical and clinical research)
- **Scopus** (broader engineering/AI and health informatics indexing)

Search strings combined controlled vocabulary and free-text terms. Core concepts included:

- ("artificial intelligence" OR "machine learning" OR "deep learning")
- AND ("diagnosis" OR "diagnostic performance" OR "screening" OR "classification")
- AND domain filters (radiology OR ophthalmology OR pathology OR cardiology OR dermatology OR "electronic health record")

ELIGIBILITY CRITERIA

Inclusion criteria

1. Peer-reviewed studies and high-quality secondary evidence (systematic reviews/meta-analyses) focused on AI-assisted **diagnostic** tasks.
2. Reported quantitative metrics (e.g., sensitivity, specificity, AUROC, accuracy) or clinically meaningful outcomes (e.g., change in detection rate, recall rate, time-to-triage).
3. Human health data (imaging, waveforms, lab/EHR, digital pathology).

Exclusion criteria

1. Non-diagnostic AI (purely administrative, scheduling, billing).
2. Purely technical papers without clinical evaluation or without performance metrics.
3. Commentary/editorials without empirical data (unless used only for context in implementation/regulation mapping).

Screening and data extraction

Titles/abstracts were screened for relevance to diagnostic AI. Full texts were examined for:

- Clinical domain and intended use
- Study type (retrospective, prospective, RCT, external validation study, meta-analysis)
- Dataset characteristics (single center vs multicenter; geographic scope)
- Diagnostic performance metrics and comparator (clinician vs AI vs combined)
- Implementation and safety themes (bias, calibration, drift monitoring, interpretability)

Data synthesis

Results were synthesized descriptively. Evidence was summarized into four structured tables:

1. Evidence landscape and study types
2. Clinical domains and common diagnostic tasks
3. Representative performance ranges from higher-quality syntheses
4. Implementation and governance themes

RESULTS

Findings – Table 1: Evidence Landscape and Validation Maturity

The evidence landscape demonstrates that most diagnostic AI studies remain retrospective validation analyses using single-center datasets, enabling rapid model development but limiting generalizability. Systematic reviews and meta-analyses provide pooled diagnostic performance estimates; however, heterogeneity in study design and ground-truth labeling is frequently reported. Prospective and randomized clinical evaluations remain comparatively scarce, reflecting logistical and cost barriers. Regulatory registry analyses confirm expanding authorization of AI diagnostic devices, though they provide limited evidence on real-world clinical effectiveness or long-term safety monitoring.

Findings – Table 2: Diagnostic Domains and AI Tasks

AI applications in medical diagnostics are predominantly concentrated in imaging-based specialties. Radiology and ophthalmology lead adoption, with AI tools performing detection, segmentation, and triage tasks to improve screening efficiency and diagnostic sensitivity. Pathology applications focus on tumor detection and grading, addressing workload and inter-observer variability. Cardiology and emergency medicine leverage AI for waveform and multimodal data interpretation to enable rapid risk stratification. Across domains, AI primarily functions as decision-support rather than autonomous diagnostic systems.

Findings – Table 3: Diagnostic Performance Ranges

Higher-quality syntheses consistently report strong discriminative performance of AI systems in narrowly defined diagnostic tasks under controlled validation settings. Mammography and diabetic retinopathy screening models demonstrate high sensitivity and specificity in curated datasets, indicating robust potential for population screening support. COVID-19 chest CT classification studies reported high pooled sensitivity, though concerns regarding dataset bias and non-representative sampling were noted. Regulatory authorization trends confirm rapid device growth, but authorization alone does not establish proven clinical outcome benefit.

Findings – Table 4: Implementation and Governance Themes

Common implementation challenges include dataset shift, where model performance declines across new populations or imaging protocols, necessitating external validation and continuous monitoring. Bias and fairness concerns arise from unequal subgroup performance, raising safety and equity implications. Variability in ground-truth labeling affects reproducibility and reported accuracy. Workflow integration determines whether AI improves or disrupts clinical practice. Finally, transparent reporting and post-market surveillance are recognized as essential to ensure accountability, regulatory compliance, and sustained diagnostic reliability.

Table 1. Evidence landscape (study design and validation maturity)

Evidence category	Typical study design	Common data sources	Strengths	Limitations frequently noted
Systematic reviews/meta-analyses	Secondary synthesis	Aggregated primary studies	Performance ranges; subgroup analysis	Heterogeneity; publication bias; inconsistent ground truth
Retrospective validation	Hold-out test sets	Single-center imaging/EHR	Rapid development; large N possible	Limited generalizability; dataset shift
Prospective observational	Live or near-live workflow	PACS/EHR integration	Real-world feasibility	Operational confounding; integration challenges
Randomized/controlled evaluations	RCTs or quasi-experimental	Clinical workflow endpoints	Stronger causal inference	Costly; complex; limited availability
Regulatory/registry analyses	Device authorization datasets	Regulatory databases	Adoption trends; risk pathways	Limited clinical effectiveness data

Table 2. Major diagnostic domains and typical AI tasks

Domain	Modalities/data	Typical AI diagnostic task	Clinical value proposition
Radiology	CT/MRI/X-	Detection, segmentation,	Worklist prioritization,

	ray/mammography	triage	sensitivity gains
Ophthalmology	Fundus photos, OCT	DR screening, referral triage	Population screening scalability
Pathology	Whole-slide images	Tumor detection/grading	Throughput, consistency, decision support
Cardiology	ECG, echo, EHR	Arrhythmia detection, risk prediction	Early detection, risk stratification
Emergency/acute care	Imaging + vitals + labs	Rapid rule-in/rule-out	Time-critical decision support

Table 3. Representative diagnostic performance ranges reported in higher-quality syntheses (illustrative summary)

Use case	Typical metric reported	Reported range (higher-quality studies)	Frequent caveat
Mammography AI assistance	AUROC / sensitivity / specificity	Often high discrimination in curated datasets	Performance varies by site, prevalence, and workflow
Diabetic retinopathy screening	Sensitivity/specificity	High values in controlled validation settings	Camera type and population differences affect calibration
Chest CT COVID-19 classification	Sensitivity/specificity	High pooled sensitivity reported in some meta-analyses	Risk of bias, non-representative datasets, label noise
FDA-authorized device trend	Count of authorized devices/year	Rapid growth, concentrated in radiology	Authorization ≠ clinical effectiveness

Table 4. Implementation, safety, and governance themes (most frequently reported)

Theme	What it means clinically	Why it matters
Dataset shift & drift	Performance drops when population/scanner/protocol changes	Requires external validation and monitoring
Bias & fairness	Unequal errors across subgroups	Safety, equity, regulatory risk
Ground truth quality	Labels may be imperfect (reader disagreement, imperfect tests)	Inflates/deflates "accuracy" and harms reproducibility
Workflow integration	AI must fit clinical pathways (triage rules, human override)	Determines whether benefit is realized
Transparency & reporting	Clear intended use, dataset description, metrics, failure modes	Enables appraisal and safe adoption
Post-market surveillance	Continuous performance tracking and updates	Required as models and environments evolve

DISCUSSION

The database synthesis indicates that **diagnostic AI performs best when the task is narrow, labels are reliable, and the evaluation setting matches the deployment setting**. This is most evident in imaging-based specialties where inputs are standardized and labels can be anchored to pathology, longitudinal outcomes, or consensus expert reading. In breast imaging, multiple studies and reviews describe meaningful opportunities for AI to support screening by improving detection and reducing workload, but also stress that benefits are conditional on careful workflow design and validation across diverse populations and devices. [11]

A recurring methodological issue is the gap between **retrospective performance** and **clinical effectiveness**. High AUROC in a benchmark dataset may not translate to fewer missed cancers, fewer unnecessary biopsies, or faster definitive diagnosis. Reasons include spectrum bias (training on "clean" cases), non-representative prevalence, and differences in image acquisition or reporting standards. In COVID-19 chest imaging, for example, some meta-analyses report high sensitivity for CT or high performance for DL models, yet broader clinical utility depended heavily on pretest probability, confirmatory testing pathways, and evolving variants and protocols. [12,13] These findings reinforce that diagnostic AI should be interpreted as a probabilistic aid, not a definitive arbiter, unless validated for a well-defined intended use.

Regulatory maturation is an important enabler of trustworthy adoption. The FDA's continuously updated list of AI-enabled medical devices demonstrates sustained growth in authorizations, particularly in radiology, and underscores that many tools are now positioned as clinical decision support or triage aids rather than autonomous diagnostic systems. [14] Year-specific analyses of FDA-authorized ML-enabled devices provide additional

granularity, describing the dominance of certain regulatory pathways and device classes, but also highlighting a persistent limitation: authorization datasets rarely prove net clinical benefit (e.g., improved outcomes) without complementary clinical studies. [15]

In Europe, governance is increasingly shaped by the EU AI Act, which entered into force on August 1, 2024, and is designed to apply progressively, with special timelines for AI embedded in regulated products. [16] For healthcare organizations and manufacturers, this implies a transition from "model development" to "lifecycle management," including risk management, transparency obligations, and documentation of robustness and bias controls. Peer-reviewed analyses discussing the AI Act in medicine emphasize compliance challenges but also clarify that high-risk healthcare AI will face stricter expectations around safety, accountability, and oversight. [17] In practice, this aligns with a broader movement toward continuous monitoring, auditability, and explicit articulation of intended use and failure modes.

Another key insight from the evidence base is that **reporting quality is not a cosmetic issue; it is a safety issue**. The CONSORT-AI and SPIRIT-AI extensions were developed to address the unique reporting needs of AI interventions, such as describing how input data are acquired, how humans interact with AI outputs, and how errors and performance are measured across settings. [18,19] More recent reporting guidance (e.g., TRIPOD+AI for prediction models using ML) extends this principle to diagnostic and prognostic models, emphasizing transparency to reduce avoidable bias and improve reproducibility. [20] For diagnostic AI specifically, transparent reporting supports informed procurement decisions, helps clinicians understand boundary conditions, and makes post-deployment performance surveillance feasible.

From an implementation standpoint, the strongest near-term value propositions are:

1. **Triage and prioritization** in high-volume imaging (reducing time-to-read for urgent cases),
2. **Second-reader augmentation** in screening settings (supporting sensitivity while maintaining specificity),
3. **Standardization** of measurements (reducing inter-reader variability for defined tasks).

However, realizing benefit requires governance infrastructure: dataset governance (including subgroup audits), technical monitoring (for drift), and clinical workflow policies (override, escalation, documentation). Without these, AI can increase noise, create over-reliance, or amplify inequities.

CONCLUSION

This database research indicates that AI has a substantial and growing role in medical diagnostics, with the most mature evidence in imaging and screening workflows. Across domains, AI demonstrates high performance for narrowly specified tasks under controlled conditions, but translation to real-world clinical benefit depends on external validation, rigorous reporting, workflow integration, and ongoing performance monitoring. Regulatory and governance frameworks—exemplified by the FDA's expanding device list and the EU AI Act's risk-based approach—are moving diagnostic AI from experimental deployment to lifecycle-managed medical technology. For clinicians and health systems, the key operational takeaway is to evaluate AI tools by intended use, validation breadth, subgroup safety, and post-deployment monitoring plans, rather than by headline accuracy alone. Future research should prioritize prospective clinical evaluations, multicenter generalizability, and measurable patient-centered impact.

REFERENCES

a. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410. doi:10.1001/jama.2016.17216. PMID:27898976.

b. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89–94. doi:10.1038/s41586-019-1799-6. PMID:31827294.

2. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph*. 2007;31(4–5):198–211. doi:10.1016/j.compmedimag.2007.02.002. PMID:17481936.

3. Kim H, Hong H, Yoon SH. Diagnostic performance of CT and RT-PCR for COVID-19: A meta-analysis. *Radiology*. 2020;296(3):E145–E155. doi:10.1148/radiol.2020201343. PMID:32301646.

4. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell*. 2021;3:199–217. doi:10.1038/s42256-021-00307-0.

5. U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Enabled Medical Devices. FDA website. Updated 2024. (Official FDA public database listing authorized AI medical devices)

6. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3:118. doi:10.1038/s41746-020-00324-0. PMID:33005057.

7. European Parliament and Council. Regulation (EU) 2024/1689 — Artificial Intelligence Act. Official Journal of the European Union. Entered into force August 2024.

8. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, CONSORT-AI Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: CONSORT-AI extension. *Nat Med*. 2020;26(9):1364–1374. doi:10.1038/s41591-020-1034-x. PMID:32807957.

9. Rivera SC, Liu X, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI Group. SPIRIT-AI extension: Guidance for clinical trial protocols involving artificial intelligence. *BMJ*. 2020;370:m3210. doi:10.1136/bmj.m3210. PMID:32873590.
10. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015;175(11):1828–1837. doi:10.1001/jamainternmed.2015.5231. PMID:26457678.
11. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of COVID-19: A review. *JAMA*. 2020;324(8):782–793. doi:10.1001/jama.2020.12839. PMID:32648899.
12. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328. PMID:32265220.
13. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: Ethical and regulatory issues in Europe and the United States. *Insights Imaging*. 2018;9(5):745–753. doi:10.1007/s13244-018-0645-y. PMID:30105404.
14. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021;27(4):582–584. doi:10.1038/s41591-021-01273-5. PMID:33753938.
15. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med*. 2020;3:126. doi:10.1038/s41746-020-00333-z. PMID:33005066.
16. Cabitza F, Campagner A, Balsano C. AI in healthcare: Risk classification, EU AI Act, and clinical implications. *Lancet Digit Health*. 2023;5(10):e698–e700. doi:10.1016/S2589-7500(23)00190-2.
17. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against healthcare professionals in medical imaging: Systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271–e297. doi:10.1016/S2589-7500(19)30123-2.
18. Park SH, Han K. Methodologic guide for evaluating clinical performance of artificial intelligence models for medical diagnosis and prediction. *Radiology*. 2018;286(3):800–809. doi:10.1148/radiol.2017171920. PMID:29319852.
19. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models (TRIPOD-AI): Need for systematic and transparent reporting. *BMJ*. 2019;367:l5501. doi:10.1136/bmj.l5501. PMID:31690501.