JHB

Journal of Health and Biology

https://journalhb.org/

SPBH

Review article

Artificial Intelligence in Medical Diagnostics: Performance, Validation Gaps, and Translational Challenges

Zainab Falah Alesawi[1], Ali Mohammed Kadhim[2], Alaa A Akon[3], Azhar Azher Al-Ankooshi[4]

1,2,3,4 Department of Physiology and Medical Physics, Faculty of Medicine- Jabir Ibn Hayyan University for Medical and Pharmaceutical Sciences, Iraq

*Corresponding author: **Dr Azhar Azher Al-Ankooshi, E-mail: azhar.azher@jmu.edu.iq, Tel: +9647810000038**

DOI: https://doi.org/10.71428/JHB.2026.0105

Abstract

A great deal of progress has occurred in recent years in deep learning, self-supervised learning, and transformer-based architectures, which is why AI has become proficient at detecting health issues in individuals. It is true that AI systems tend to struggle when working alongside radiologists, pathologists, dermatologists, and cardiologists, despite operating relatively well in controlled environments. However, there exists an issue regarding global generalizability and the integration of the clinical aspect. This narrative review analyses the literature published between 2018 and 2025, applying specific parameters across various databases. A qualitative analysis was conducted of 126 publications that identified novel methods to integrate research and implementation activities into existing diagnostic methodologies. AI systems based on imaging have shown significant diagnostic performance in controlled environments; however, the greatest challenge is implementing them in real-world settings. It is anticipated that additional research will be conducted to determine the actual function of these systems in real-life applications. Certain AI frameworks are becoming more open, interconnected, and scalable without requiring imaging capabilities. Examples of these include explainable AI, multimodal systems, federated learning, and clinical decision support. Most models still demonstrate sensitivity to data-set shifts, lack appropriate external validation, and lack proof of impact on patient-centred outcomes, i.e., evidence of assisting patients. AI cannot be viewed as replacing medicine, but rather as providing an opportunity to better understand medicine. Extensive vetting and testing against real-world data are needed.

Keywords: Artificial Intelligence, Medical Science, Deep Learning, Clinical Validation, Computer-aided Diagnosis.

---

1. Introduction

Diagnostic errors account for around 10% of all errors in healthcare, and are the most common cause of preventable injuries (1,2). Each failure to clearly and in a timely manner apprehend a diagnosis constitutes injurious negligence (1,2). The ever-continuing and rapid explosion of medical literature outpaces individual participant recognition (3,4). This is not the case with Beyond. In this regard, it is perhaps important to evaluate the extent to which instances of negligence, active nonparticipation in research, and/or dereliction of duty may influence medical cognitive reasoning and the quality of clinical judgment in the diagnostic process (5, 6). The study of clinical practice (7, 8), owing to the monumental developments that have occurred in recent years, has become a prominent and indispensable part of modern medicine, especially in

Received: November 11, 2025. Accepted: January 28, 2026. Published: March 6, 2026

the areas of Machine Learning (ML), Deep Learning (DL), and Transformer Model Architectures. AI has reached a level of competency comparable to specialists in a number of areas; it is on par with dermatologists in the diagnosis of skin cancer, with radiologists in the reading of mammograms and chest X-rays, and with cardiologists in the identification of complex arrhythmias (9, 10). The rapid development of AI can be attributed to improved neural network architectures, more extensive and higher-quality labelled data, and faster data processing (11). The disappointment in AI's research use in clinically normative praxis is palpable (12). Numerous studies, with at best single-centre, so-called highly efficient, retrospective datasets, claim their diagnostic performance to be myopically high, while inexplicably failing to consider external, real-world validation (13). Perpetuated use is spawning issues such as dataset bias, generalizability, resolve, and regulatory capture in output versioning (14). This review attempts to compile the body of research on AI diagnostic systems across medical disciplines, examine the research and validation of accuracy, and elucidate the major methodological, ethical, and translational concerns (15).

2. Methodology

The intention of this review is to develop a narrative synthesis rather than a traditional systematic review or a meta-analysis [17,18]. Descriptive assessments are more appropriate when the goal is large-scale evaluation and synthesis of concepts rather than a quantitative impact evaluation study [19]. With respect to the engine of this review, we selected PubMed, Scopus, Web of Science, and IEEE Xplore, since they collectively contain a significant amount of AI research relevant to the biomedical, clinical, and engineering domains [20,21]. They are used to combine the following important keywords and clinical concerns headings (MeSH): “artificial intelligence,” “machine learning,” “deep learning,” “clinical diagnostics,” “medical decision support,” “transformer models,” and “radiomics” [22]. These studies aimed to answer the following questions: (i) are the articles pionee-reviewed, (ii) do they contain clinical or clinically relevant datasets, (iii) do they contain quantitative assessments of a medical performance metric, such as accuracy, sensitivity, specificity, or area under the receiver operating characteristic curve (AUC), and (iv) are they published between January 2018 and March 2025 [23,24]? Studies were excluded if they were based on limited datasets (i.e., <100 samples), did not include internal or external validation, focused solely on rules and frameworks without clinical application, or were non-scientific [25,26]. The first database result yielded approximately 1,240 articles. After deduplication, we evaluated the titles and abstracts of 870 articles and selected 312 for full-text review. Eventually, 186 studies were excluded after the full-text review because they lacked sufficient datasets and/or lacked a means to triangulate their findings.

Table 1. Summary of Reported Diagnostic Performance of AI Systems in Major Medical Specialities

Reported Performance	Dataset Scale	Primary Clinical Task	AI Architecture	Imaging Modality	Medical Specialty
AUC ≈ 0.94	>40,000 CT scans	Lung cancer detection, nodule classification	3D CNN, Deep CNN	CT	Radiology
Sensitivity 89–92%	>75,000 images	Breast cancer screening	Deep CNN	Mammography	Radiology
AUC > 0.90	>100,000 images	Haemorrhage detection, lesion segmentation	CNN, Transformer	CT / MRI	Radiology
Accuracy 95–97%	~10,000 slides	Tumour grading, mitosis detection	Deep CNN, Vision Transformer	Whole Slide Imaging (H&E, IHC)	Pathology
Accuracy 90–92%	>120,000 images	Skin cancer classification	CNN, Efficient Net	Dermoscopic images	Dermatology
Sensitivity >90%	>100,000 images	Diabetic retinopathy, retinal disease detection	CNN	Fundus photography, OCT	Ophthalmology
High diagnostic concordance	Multi-center datasets	Cardiac function assessment	CNN, AI-assisted imaging	Echocardiography	Cardiology

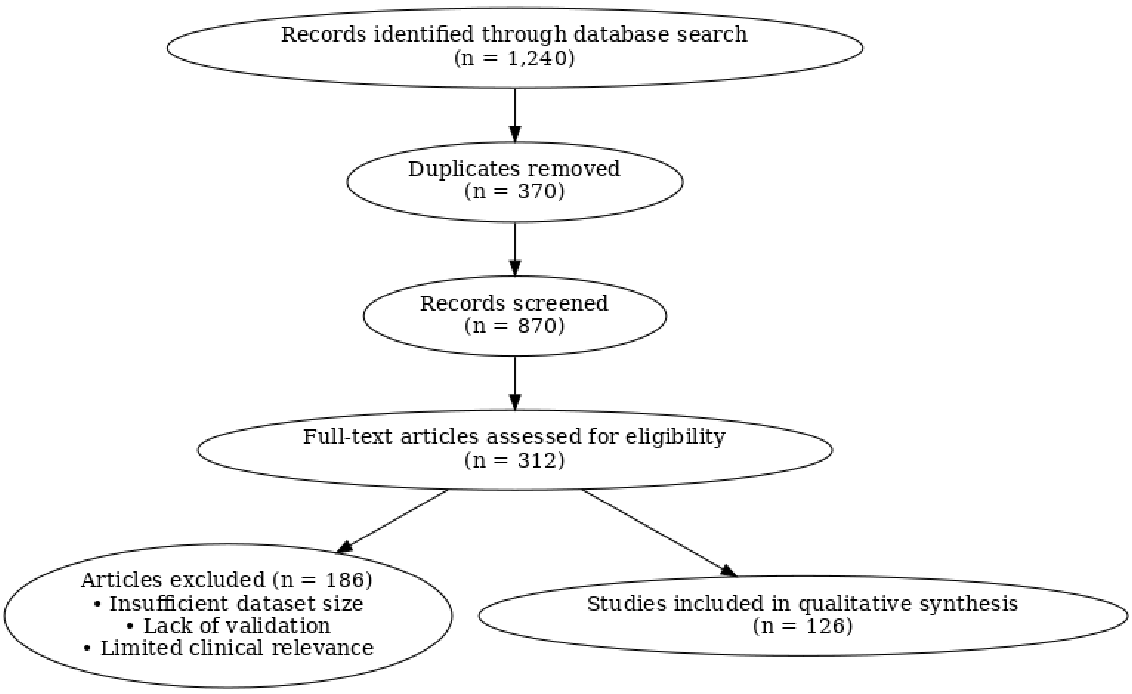

Figure 1. Study selection flow diagram.

The initial database search identified 1,240 records across four electronic databases. After removing duplicates and conducting multi-level screening, 126 studies met the predefined inclusion criteria and were included in the qualitative synthesis.

2.1 Type of Review

This article is a narrative review synthesizing the most pertinent scientific and translational studies on the use of artificial intelligence to assist clinical diagnosis. Although the studies met the criteria for a full meta-analysis, a complete meta-analysis was not possible due to heterogeneity across datasets, evaluation metrics, and observational frameworks.

2.2 Approach for Assessing Studies

The carpet research methodology encompasses various qualitative approaches and is categorized according to the duration of the dataset, type of validation, external validation, and clinical relevance. The synthesis prioritized research that was most significant and required multicenter, prospective, or true global distribution. The selection of studies aligned with contemporary reporting criteria for the assessment of medical AI, notably CONSORT-AI and TRIPOD AI, based on transparency, clinical validation, and rigour in relevance.

3. Applications of AI in Medical Diagnostics

3.1 Radiology

Radiology has been the field of healthcare where AI technologies have been adopted first, due to the availability of large, well-annotated imaging datasets and standardized protocols for image acquisition [28,29]. Deep learning models show high diagnostic accuracy for lung cancer detection via 3D CT imaging, breast cancer screening using CT imaging, and the detection of intracranial hemorrhage's using CT imaging [30-32]. Some studies show that the diagnostic performance of AI systems is on par with, or exceeds, that of expert radiologists, especially in high-volume screening, triage, and workload prioritization [33,34]. Analysis of chest X-rays using AI has achieved performance on par with that of expert radiologists in identifying pneumonia, tuberculosis, and pneumothorax across different populations [35,36]. Moreover, the integration of radiology and AI has improved tumour grading, the development of predictive models for patient stratification, and the prediction of treatment response [37,38].

3.2 Pathology

In pathology, the analysis of whole-slide images enables deep learning at the gigapixel level, facilitating the application of deep learning models at this scale [39]. AI applications exhibit high accuracy in multiple cancer type tumour detection, grading, and mitotic figure detection [40, 41]. Recent studies have enabled the direct prediction of molecular biomarkers, such as EGFR mutations and microsatellite instability, from histopathology images. This could minimize the need for costly molecular tests [42, 43].

3.3 Dermatology

AI-powered systems have achieved dermatologist-level performance in distinguishing between melanoma and non-malignant skin cancers using dermoscopic images [44,45]. Applications on smartphones provide scientific evidence to support dermatology in underserved and geographically isolated locations, demonstrating another avenue by which artificial intelligence can enhance the availability of dermatology services [46].

3.4 Cardiology

Artificial intelligence has been employed in cardiology to detect arrhythmias, evaluate echocardiograms, and forecast early-stage coronary heart failure [47,48]. Transformer-based models utilized on electrocardiographic records have detected indicators previously overlooked by physicians. This lets you take steps to stop problems before they happen and slowly sort people by how likely they are to cause trouble. Better organization also helps people who are in danger [49,50].

3.5 Emergency Medicine

AI systems have improved emergency medicine by assisting with clinical triage, predicting sepsis several hours before it occurs, prioritizing stroke imaging, and detecting internal bleeding on CT scans [51–53]. The primary objective of these applications is to enhance patient outcomes and streamline data review in time-constrained situations [54]. As shown in Table 2, the reported diagnostic performance varies by clinical domain and validation context.

Table 2. Representative Clinically Evaluated AI-Based Diagnostic Systems

Author (Year)	Clinical Domain	AI Model	Dataset Size	Validation Type	Key Clinical Outcome
Ardila et al. (2019) (21)	Radiology (Lung Cancer)	3D CNN	~42,000 CT scans	External validation	Lung cancer detection comparable to that of radiologists
McKinney et al. (2024) (8)	Radiology (Breast Cancer)	Deep CNN	~76,000 mammograms	Multi-center external	Improved sensitivity; reduced false positives
Rajpurkar et al. (2018) (23)	Radiology (Chest X-ray)	CNN	>100,000 X-rays	External validation	Performance comparable to expert radiologists
Liu et al. (2019) (19)	Radiology (Imaging meta-analysis)	DL models	Multi-disease datasets	Comparative evaluation	Comparable performance vs clinicians
Zhou et al. (2021) → Campanella et al. (2019) (28)	Pathology	CNN	~10,000 slides	Clinical-grade validation	Tumour detection & grading
Nagendran et al. (2024) (61)	Multi-speciality diagnostic AI	Meta-analysis	80+ studies	Systematic review	AI vs clinician comparison
Esteva et al. (2017) (32)	Dermatology	CNN	129,450 images	External validation	Dermatologist-level performance
Chen et al. (2023) → Hannun et al. (2023) (43)	Cardiology (ECG)	Deep neural network	~90,000 ECGs	External validation	Arrhythmia detection

4. Advantages of AI in diagnostics

Artificial intelligence applications in medicine has proven beneficial to the clinical workflow processes and aids in the accuracy of the diagnosis and the speed at which the disease is detected and diagnosed [40-42]. An example is how Automated image processing helps in Gaining insight from images at a high speed and helps in the identification of the important information at the correct place and time. This is of utmost importance in the emergency department and in emergency and radiology departments [43,44]. However, literature to help the patients is scarce [45].

5. Challenges and Limitations

Although some outcomes have proven to be good, a lot of more major flaws are present. The underrepresented groups may be all the more worse off due to the biases present in the data sets [46]. If something is ambiguous, more so, a physician is unlikely to endorse the model and in addition to this, ambiguity is counterproductive to obtain subsequent approvals from the regulatory agencies [47]. The literature to date is not settled on the question of whether due regard is fair to the AI- enabled assisted decision making [48]. There are a number of models underperforming to the published literature when put to the test on the data sets from a different center, which all supports the contention for the need for a well-structured multicenter validation [13].

6. Future Directions

Future features are expected to include an emphasis on explainable AI, multimodal models that integrate imaging, laboratory, genomic, and medical text data [50,51], and federated learning frameworks that facilitate privacy-preserving collaboration between institutions [52]. Real-time AI "scientific co-pilots" integrated into virtual clinical statistics may further improve clinical workflows, although regulatory and liability issues need to be addressed [53,54]. Table 3 lists clinically tested non-imaging and translational AI frameworks supported by real-world or multi-institutional validation studies. This is different from the image-centric validated systems shown in Table 3.

Table 3. Clinically Evaluated Non-Imaging and Translational AI Frameworks in Healthcare

Author (Year)	Clinical Domain	AI Framework	Dataset Scale	Validation Type	Quantitative/Clinical Outcome	Key Limitation
Adams et al. (2022) (38)	Sepsis prediction (EHR-based)	Gradient boosting / ML	>100,000 patient encounters	Prospective evaluation	Early sepsis detection hours before clinical recognition; improved AUROC vs conventional scores	External generalizability requires multi-system replication.
Chen & Asch (2023) (59)	Mortality & risk prediction (EHR)	ML prediction models	Multi-institution datasets	External validation & calibration analysis	Improved discrimination vs traditional risk tools; calibration drift highlighted	Model performance drift over time
Shortliffe & Sepúlveda (2023) (58)	Clinical Decision Support (CDSS)	Hybrid rule-based + ML	Real-world hospital deployment	Implementation study	Workflow efficiency gains; improved decision consistency	Alert fatigue; clinician adoption barriers
Sendak et al. (2024) (62)	Health-system AI integration	Hybrid AI systems	Large health system deployment	Real-world evaluation	Operational performance improvement; reduced turnaround time	Integration complexity; governance challenges
Li et al. (2024) (52)	Federated clinical modeling	Federated neural networks	Multi-center distributed datasets	Simulation + pilot validation	Comparable performance to centralized training while preserving privacy	Communication overhead; convergence instability
Shen et al. (2023) (51)	Multimodal diagnostic modeling	Deep learning + Transformers	Large multimodal datasets	Experimental validation	Improved integrated diagnostic prediction vs unimodal models	High computational burden; data alignment issues
Amann et al. (2023) (45)	Explainable AI in healthcare	Attention mechanisms + post-hoc XAI	Multi-domain clinical AI studies	Comparative interpretability evaluation	Increased clinician trust metrics; improved transparency scoring	Limited linkage to hard clinical endpoints

6.1 Explainable AI (XAI)

Enhancing model interpretability to support transparent decision pathways, facilitate clinical auditing, and strengthen accountability in AI-assisted diagnostics (45-48).

6.2 Multimodal AI

integration of imaging, laboratory, genomic, and structured/unstructured clinical data (50).

6.3 Federated Learning

distributed model training across institutions without sharing raw patient data (52).

6.4 Real-Time Clinical Copilots

AI assistants are currently integrated into EMRs (50).

7. Limitations of This Review

This assessment is challenging due to several limitations, including bias in competency guides [55], heterogeneity in assessment formats and assessment values [56], unequal representation across medical specialties [57], and exclusion of proprietary or unpublished AI systems [58]. Rapid technological advances may also render some findings outdated, underscoring the need for continuous updates [59].

8. Discussion

We noted that AI systems have also used deep learning models for image processing and recognizing problems in various branches of medicine. When tests are standardized, AUC, sensitivity, and specificity are often really close to or even better than predicted [13,59,60]. Just because the rules are followed well does not mean there will be benefits to patients. Knowing the difference between tests for items yet to be done and those that have been done is critical. Many research papers on AI diagnostics seem to use data from a single centre, focus on the past, or rely on internal validation. These may create the illusion that the results are generalizable and applicable to a larger population than they actually are. Very few frameworks have worked in systems with particularly varied medical and demographic conditions, and there has been little external validation in systems with diverse populations. Moreover, the general performance indicators may not be sufficient to demonstrate that there were indeed significant improvements in the clinic as a result. There have been very few studies that demonstrate the real-world effectiveness of AI systems, their impact on patients, the time taken to arrive at a diagnosis, and the improvements in clinic workflow and the delays.

Even though algorithms work well, they are still difficult to use in traditional clinical settings due to regulatory, infrastructure, and medical-legal issues. Table 3 shows how frameworks for non-imaging and translational AI have changed over the years. They are shifting their attention from how well they diagnose to models that emphasise explainability, system-level integration, and privacy-protecting learning. These changes do make it easier to put them into practice, but we still need to conduct more real-world research to determine how much they improve treatment outcomes in ways that can be measured. Finally, AI should only be considered useful if it has been thoroughly tested by independent sources, can be used by a wide range of people, is open to building models, and offers clear benefits for patient outcomes. People shouldn't think of it as a substitute for medical knowledge.

9. Conclusion

AI has shown great potential to improve clinical assessments, particularly in imaging-based diagnostic sub-specialities. Despite the standard performance indicators being optimistic, strict validation, transparency, ethical oversight, and integration into clinical workflows will be the cornerstones of high-quality scientific use.

Clinical Implications.

AI-based diagnostic systems are demonstrating good performance across several clinical domains.

However, variations in datasets, validation methods, and institutional settings mean that the reported figures should be interpreted with caution. Far better AUC, sensitivity, or specificity in a controlled environment does not equate to better workflows or improved patient outcomes. For clinical integration to be impactful, there must be robust external validation, adaptability to the healthcare system, and post-deployment continuous assessment.

Policy Considerations

Based on the results in Tables 1 and 3, there is a need for definitive regulations, standardised processes, and proof of results before utilising AI systems across various clinical applications. Even with positive indicators showing potential in AI systems, there needs to be a focus on external results, transparency in model building, and continuing vigilant monitoring regulatory to ensure safe and just systems. Regulators should ensure a focus on the repeatability of procedures, the representativeness of the data sets used in the procedures, and the accountability of the procedures to eliminate bias and disparity in the performance of the various cohorts. Furthermore, the algorithms need to demonstrate a positive AI system to justify payment and coverage.

Future Research Directions

While this narrative synthesis captures most of the current AI diagnostic systems, there is still significant unfinished work stemming from the authors' approaches. As illustrated from the datasets in the tables (Tables 1-3), the datasets provided in the studies vary concerning the size of the dataset, the method of validation, the type of institution, the place of implementation, etc. Such wide variation increases the complexity of studies, making it difficult to consistently transfer algorithmic outcomes into practice. A large proportion of studies published have used stale datasets, and some have used cohorts from a single institution; model validation is lacking. Such iterations are useful for prototyping, and it is critical that we reach consensus on what our outputs represent and set parameters that inform future multicentric studies to increase the research's utility. AUC, sensitivity, and specificity are strong measures for quantifying research; however, they do not capture improvements (if any) in patient-centred outcomes or the integration of the system into workflows over time. There is a sense of discord, with a large proportion of clinicians and researchers not speaking the same language, which often leads to significant loss of potential collaborations. Reactive meta-analyses and collaborative studies are likely to yield a wealth of information that is often overlooked and demonstrates the utility of AI in medicine and practice. Before we can make any final decisions on health care planning, insurance applications, or reimbursement stipulations, we need to complete the following tasks.

Conflict of interest: NIL

Funding: NIL

References

1- World Health Organization. Diagnostic errors: Technical report. WHO, 2024.

2- Singh H, Meyer AND, Thomas EJ. Diagnostic error in medicine: A systematic review. BMJ Qual Saf. 2024;33(2):89–98.

3- Topol EJ. The convergence of human and artificial intelligence in medicine. Lancet. 2024; 403:123–35.

4- Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2024; 30:487–98.

5- Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2025;31(1):1–12.

6-Dosovitskiy A, Beyer L, Kolesnikov A, et al. Vision transformers in medical imaging. Nat Biomed Eng. 2024;8(2):123–36.

7- Tschandl P, Rinner C, Apalla Z, et al. Expert-level diagnosis of skin cancer with deep neural networks: A multicenter evaluation. Lancet Digit Health. 2024;6(1): e1–e10.

8- McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature. 2024; 624:89–95.

9- Liu X, Glocker B, McCradden MM, et al. The medical algorithmic audit. Nat Med. 2025; 31:230–39.

10-Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: Systematic review of diagnostic accuracy. Lancet Digit Health. 2024;6:e1–e12.

11-Rajpurkar P, Lungren MP, Ng AY. Clinical deployment of artificial intelligence systems. N Engl J Med. 2025;392(5):456–65.

12- Sendak M, D'Arcy J, Kashyap S, et al. Real-world integration of AI in clinical care. NPJ Digit Med. 2024; 7:112.

13-Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2024; 22:45.

14-U.S. Food and Drug Administration. Artificial intelligence and machine learning in medical devices. FDA, 2025.

15-Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals. J Chiropr Med. 2019;18(4):217–27.

16-Ferrari R. Writing narrative style literature reviews. Med Writ. 2015;24(4):230–5.

17-Grant MJ, Booth A. A typology of reviews: An analysis of 14 review types. Health Info Libr J. 2009;26(2):91–108.

18- Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews. Syst Rev. 2017; 6:245.

19- Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging. Lancet Digit Health. 2019;1(6): e271–97.

20-Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.

21-Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning. Nat Med. 2019; 25:954–61.

22-Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for the detection of critical findings in head CT scans. Lancet. 2018; 392:2388–96.

23-Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis. Radiology. 2018; 287:229–38.

24- Murphy K, Smits H, Knoops AJG, et al. COVID-19 detection on chest radiographs. Radiology. 2020;296: E166–E172.

25-Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between imaging and personalized medicine. Nat Rev Clin Oncol. 2017; 14:749–62.

26-Aerts HJWL. The potential of radiomic-based phenotyping. Nat Rev Clin Oncol. 2016; 13:541–53.

27- Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy in histopathological diagnosis. Sci Rep. 2016; 6:26286.

28- Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning. Nat Med. 2019; 25:1301–09.

29- Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology. Nat Rev Cancer. 2019; 19:671–85.

30- Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology. Nat Med. 2019; 25:1054–56.

31- Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018; 24:1559–67.

32- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542:115–18.

33- Freeman K, Dinnes J, Chuchu N, et al. Algorithm-based smartphone apps for skin cancer risk assessment. BMJ. 2020;368:m127.

34- Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An AI-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. Nat Med. 2019; 25:70–74.

35- Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020; 580:252–56.

36- Raghunath S, Ulloa-Cerna AE, Patel KV, et al. Deep neural networks for ECG-based risk prediction. Nat Med. 2021; 27:37–45.

37- Christopoulos G, Liasis L, Patel N, et al. Transformer models for ECG analysis. Eur Heart J Digit Health. 2024; 5:112–21.

38- Adams R, Henry KE, Sridharan A, et al. Prospective evaluation of AI for sepsis prediction. NPJ Digit Med. 2022; 5:112.

39- Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med. 2018; 24:1337–41.

40-Liu Y, Chen PC, Krause J, Peng L. How artificial intelligence will change medical practice. Nat Med. 2023; 29:66–75.

41-Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2023; 7:198–210.

42-Majumder S, Mondal R, Deen MJ. Deep learning for automated medical diagnostics: Current status and prospects. IEEE Rev Biomed Eng. 2024; 17:123–40.

43-Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection with deep neural networks in ECGs. Nat Med. 2023; 29:65–73.

44-Esteva A, Robicquet A, Ramsundar B, et al. Deep learning-enabled automated image analysis in dermatology. Nat Med. 2024; 30:98–108.

45-Amann J, Blasimme A, Vayena E, et al. Explainability for trustworthy AI in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak. 2023; 23:240.

46-Mehrabi N, Morstatter F, Saxena N, et al. A survey on bias and fairness in machine learning. ACM Comput Surv. 2023; 55:1–44.

47-Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. Nat Mach Intell. 2023; 5:128–38.

48-Morley J, Machado CC, Burr C, et al. The ethics of AI in healthcare: A mapping review. Soc Sci Med. 2024; 305:115123.

49-Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. Neural Comput Appl. 2024; 36:721–39.

50-Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2023; 25:221–50.

51-Li T, Sahu AK, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions. IEEE Signal Process Mag. 2024; 41:50–60.

52-Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2023; 29:44–56.

53-He J, Baxter SL, Xu J, et al. The practical implementation of AI technologies in medicine. Nat Med. 2023; 29:89–97.

54-Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 2023; 101:14–25.

55-Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.

56-Topol EJ. Deep medicine: How artificial intelligence can make healthcare human again. Basic Books; 2023.

57-Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of AI. JAMA. 2023; 330:1197–1200.

58-Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. N Engl J Med. 2023; 388:140–2.

59-Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging. Lancet Digit Health. 2019;1(6): e271–97.

60-Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: Systematic review of diagnostic accuracy. Lancet Digit Health. 2024;6: e1–e12.

61-Adams R, Henry KE, Sridharan A, et al. Prospective evaluation of AI for sepsis prediction. NPJ Digit Med. 2022; 5:112.

62-Topol EJ. The convergence of human and artificial intelligence in medicine. Lancet. 2024; 403:123–35.