International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Artificial Intelligence-Driven Diagnosis in Healthcare: Applications, Benefits, Limitations, and Challenges

Deependra Pratap Singh, Divya Singh, and Devansh Singh
deependrapratapsingh7570@gmail.com, divyasingh.jd@bbdu.ac.in, singhdevansh0071@gmail.com
School of Computer Applications, Babu Banarasi Das University, Lucknow, Uttar Pradesh, India

***Abstract*—**Diagnosis is the step on which every later clinical decision rests, and it remains vulnerable to workload, fatigue and incomplete information. Artificial intelligence (AI), chiefly machine learning and deep learning, is now being applied to this step in imaging, pathology, ophthalmology, dermatology and clinical text. This paper is a theoretical review. It follows the development of AI-assisted diagnosis, sets out the main application areas, compares AI-assisted practice with conventional practice, and examines the limitations that restrict clinical use: data quality and bias, weak external validation, opacity, privacy, accountability and integration costs. Published evidence, including regulatory figures up to 2025, is used to ground each point. The paper reports no experiments. It concludes that current evidence supports AI as a decision-support aid under human supervision, and that responsible adoption depends on diverse data, prospective validation, interpretable output and clear regulation.

***Index Terms*—**AI-driven diagnosis, deep learning, explainable AI, healthcare, medical imaging, clinical decision support.

I. INTRODUCTION

Diagnosis means naming the condition that explains a patient's symptoms, history, test results and images. It is done by people, and people err. A widely cited estimate from three US observational studies puts the outpatient diagnostic error rate at about 5% of adults each year [6]. Part of the cause is cognitive: fatigue, time pressure, anchoring on a first impression. Part is structural, since a single case can involve laboratory panels, scans and free-text notes that no clinician can fully cross-check in a ten-minute consultation.

AI refers to computer systems that perform tasks normally associated with human reasoning, such as pattern recognition and prediction. Within it, machine learning (ML) fits models to labelled examples rather than to hand-written rules, and deep learning (DL) does so with multilayer neural networks that learn their own features from raw data [2], [5]. These methods have matured quickly in diagnosis because medicine produces large volumes of digital data, and because image-based specialties give a clear task: look at a picture, decide whether disease is present.

The aim here is modest. This paper reviews what AI-driven diagnosis does at present, what benefits are documented, and what holds it back, so that a beginner in the field can see the whole picture in one place. It does not propose a new algorithm and it does not run an experiment. Section II states the review approach, Section III traces the background, Section IV describes applications, Section V outlines a typical diagnostic pipeline, Sections VI and VII compare AI-assisted and conventional practice and list the benefits, Section VIII covers limitations, Section IX considers governance, and Section X closes with future directions.

II. REVIEW APPROACH

The study is secondary research. It draws on the synopsis prepared for this work and on peer-reviewed journals, regulatory publications and international guidance. Selection followed three rules. A source had to concern diagnosis, not treatment or hospital administration. It had to report an evaluation, a regulatory fact, or a clearly argued critique. Where several sources made the same point, the one with the stronger design was preferred, so that a randomised trial outweighs a retrospective comparison and a retrospective comparison outweighs an opinion piece.

The analysis then followed the stages set out in the synopsis: literature review, collection of examples from imaging, ophthalmology, dermatology and related fields, comparison with conventional practice, and identification of challenges. Figures quoted from regulators and trials are reported as the sources state them. Because registries and guidance documents are revised often, such figures describe the position at the time of writing and should be rechecked before reuse.

III. BACKGROUND AND RELATED WORK

A. From Expert Systems to Deep Learning

Early medical AI took the form of rule-based expert systems, which encoded specialist knowledge as if–then statements. They were transparent but brittle: every rule had to be written and maintained by hand, and they coped badly with the variety of real patients [5]. Statistical ML replaced them. Decision trees, support vector machines, logistic regression and random forests learn associations from records, and they remain common for tabular data such as laboratory values and risk factors [5].

The present phase began with convolutional neural networks (CNNs). Once large labelled image sets became available, CNNs matched or approached specialist performance on narrow tasks. Gulshan *et al.* trained a network to grade diabetic retinopathy from retinal photographs and reported high sensitivity and specificity

against an ophthalmologist reference [7]. Esteva *et al.* reported dermatologist-level skin lesion classification [8]. Rajpurkar *et al.* described CheXNet, a 121-layer network for pneumonia detection on chest radiographs that was compared with radiologists on a test set [3]. Topol frames these results as the start of a shift in which clinicians work alongside algorithms [1].

B. Language Models and Clinical Text

Natural language processing (NLP) extracts information from clinical notes and reports, which are largely unstructured. Large language models extend this. Singhal *et al.* showed that a model tuned for medicine reached a pass-level score on US licensing-style questions, while also documenting gaps against clinicians in the quality of answers [13]. Passing an examination is not the same as diagnosing a patient, and the authors say so. The result is better read as evidence that clinical knowledge can be encoded than as evidence of readiness for bedside use.

C. Current Scale of Adoption

Regulatory records show how far deployment has gone. The first autonomous AI diagnostic system, IDx-DR for diabetic retinopathy, received US Food and Drug Administration (FDA) authorisation in 2018 after a pivotal trial in primary care [11]. The FDA now maintains a public list of AI-enabled medical devices; by 2025 it contained well over one thousand entries, and radiology accounted for roughly three-quarters of them [12]. The distribution is uneven. Most cleared devices read images, and few address diagnosis from multimodal records.

IV. APPLICATIONS OF AI IN DIAGNOSIS

A. Medical Imaging

Imaging is the most mature domain. Systems flag suspicious regions on radiographs, CT, MRI and ultrasound, prioritise urgent studies in a worklist, and measure lesions consistently. In breast screening, McKinney *et al.* found that an AI system reduced false positives and false negatives relative to the original radiologist readings in UK and US datasets, and that it could stand in for the second reader in the UK double-reading workflow [9]. The MASAI randomised trial in Sweden later reported that AI-supported screen reading preserved cancer detection while cutting the reading workload by roughly 44% [10]. These are among the strongest findings in the field because one is a prospective trial, not a retrospective comparison.

TABLE I
REPRESENTATIVE EVIDENCE BY DIAGNOSTIC AREA

Area	Task	Evidence
Retina	Diabetic retinopathy grading	[7], [11]
Skin	Lesion classification	[8]
Chest	Pneumonia on radiographs	[3]
Breast	Screening mammography	[9], [10]
Text	Clinical question answering	[13]

B. Ophthalmology and Dermatology

Diabetic retinopathy screening suits automation. The disease is common, the retinal image is standardised, and the shortage of specialists is acute in many regions. Autonomous systems can return a referral decision at a primary-care visit without a specialist present [7], [11]. In dermatology, classifiers trained on clinical and dermoscopic photographs support triage, though their performance depends heavily on skin tone and image quality in the training set [8].

C. Pathology, Cardiology and Infectious Disease

Digitised pathology slides contain billions of pixels, which suits automated pre-screening and counting. In cardiology, models read electrocardiograms and echocardiograms to detect rhythm disorders and reduced pumping function. For infectious disease, chest-radiograph tools were widely developed during the COVID-19 pandemic, but a systematic review found that nearly all of the published prediction models were at high risk of bias and unfit for clinical use [15]. The episode is a useful caution: speed of development was not matched by quality of validation.

D. Risk Prediction and Decision Support

Beyond reading single images, ML models combine age, history, laboratory trends and genetic data to estimate disease risk, supporting prevention and prioritising high-risk patients [1], [5]. Such tools act as decision support: they raise a prediction or a prompt, and the clinician decides.

V. A TYPICAL DIAGNOSTIC PIPELINE

Most deployed systems follow the same sequence, illustrated in Fig. 1. Data are first acquired from a scanner, camera, laboratory system or record. They are then cleaned and standardised, for instance by resizing images, normalising intensity or mapping terms to a common vocabulary. A trained model produces a score, a label or a highlighted region. The output reaches a clinician, who accepts, corrects or rejects it, and the final decision is recorded together with the outcome.

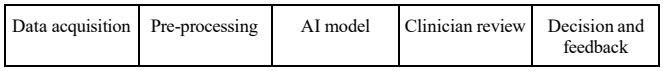


Fig. 1. Generic flow of an AI-assisted diagnostic system with clinician review.

Each stage carries its own failure modes. Acquisition varies between devices and sites. Preprocessing can discard information or introduce artefacts. The model can be wrong with high confidence. Presentation can nudge the reader toward the suggestion. The record of outcomes is what allows later monitoring, yet it is often incomplete because confirmed diagnoses arrive weeks or months afterwards. Reading the pipeline as a chain makes plain that a gain at one stage can be cancelled at another, and that evaluating the model alone says little about performance in practice [4], [24].

Two deployment patterns are common. In the first, the system acts as a second reader or concurrent aid: the clinician sees the suggestion while reading. In the second, it acts as a triage filter that sorts studies by urgency or clears obviously normal ones. The first mostly affects accuracy, the second mostly affects workload, and the evidence cited earlier for screening mammography concerns both [9], [10]. Autonomous operation, as in the diabetic retinopathy system cleared in 2018, remains limited to tasks where the input is standardised and the output is a simple referral decision [11].

VI. AI-ASSISTED VERSUS CONVENTIONAL DIAGNOSIS

Table II compares the two approaches along the factors identified in the synopsis. The comparison is qualitative and generalises across studies, so individual systems will differ.

TABLE II
COMPARISON OF CONVENTIONAL AND AI-ASSISTED DIAGNOSIS

Factor	Conventional	AI-assisted
Speed	Limited by clinician time	Rapid preliminary output
Consistency	Varies between readers and with fatigue	Same input gives same output
Data handled	Selected by clinician	Large, multimodal volumes
Context	Strong: history, patient setting	Weak outside training data
Explanation	Reasoned by clinician	Often opaque
Main risks	Error, fatigue, access	Bias, drift, overreliance

Speed and consistency favour the algorithm. Judgement in unusual cases, communication with the patient and responsibility for the decision favour the clinician. The evidence therefore points to combination, not substitution. In the MASAI trial the algorithm triaged and supported the radiologists, who kept the final decision [10]. Experience with decision aids also warns of automation bias, where a clinician accepts a wrong suggestion because it came from a machine [1], [5].

VII. DOCUMENTED BENEFITS

A. Earlier Detection

Screening programmes work only if disease is caught while it is treatable. Evidence from breast screening indicates that AI support can hold or raise detection at a lower reading burden [9], [10]. Retinal screening shows a similar pattern for diabetic eye disease, where a referral made in primary care can arrive years before symptoms would prompt one [7], [11].

B. Workload and Access

The MASAI result is notable chiefly for what it implies about capacity: if a screening service can read the same volume with far fewer reader hours, specialist time can be redirected to complex cases [10]. The same argument applies to regions with few specialists. Rural and underserved populations may gain a preliminary assessment that would otherwise not be available, provided that referral pathways and medical supervision exist behind the tool [1], [20]. A system that identifies disease nobody can then treat does not improve outcomes.

C. Consistency and Documentation

An algorithm applies the same criteria to every case and does not tire at the end of a shift. It also leaves a record of what it saw, which supports audit and teaching. Measurement tasks, such as lesion size or ejection fraction, benefit especially because human estimates vary between readers [2], [5]. These advantages are real but narrow. They concern repeatable perception tasks and do not extend to weighing a patient's circumstances, preferences and competing conditions.

VIII. LIMITATIONS AND CHALLENGES

A. Data Quality, Bias and Fairness

A model can only reflect the data it was trained on. If a dataset over-represents one region, scanner type or demographic group, accuracy on others may fall. The clearest documented case is not an imaging tool but a commercial risk algorithm that used past healthcare spending as a proxy for need; because less money had historically been spent on Black patients with the same illness burden, the algorithm under-identified their needs [16]. Labels are a related problem. Ground truth in diagnosis is often a clinician's opinion or a later test, and where no definitive standard exists, "accuracy" is partly circular [17].

B. Validation and Generalisation

Many published models are evaluated retrospectively, on data from the institution that built them. Performance often drops at a new hospital with different equipment, protocols and patients [4], [5]. Data leakage, where information from the test set seeps into training, inflates reported results across ML-based science generally [14]. Prospective multi-site trials such as MASAI remain the exception [10]. Liu *et al.* give clinicians a checklist of questions for reading such studies, among them whether the test set was truly independent [4].

C. Explainability

Deep networks do not state their reasons. Explainable AI (XAI) techniques, such as heat maps that highlight the image regions driving a prediction, try to fill this gap. Their value is disputed. Rudin argues that for high-stakes decisions inherently interpretable models should replace post-hoc explanations of black boxes [18]. Ghassemi *et al.* add that current heat maps can be unreliable and may give clinicians false reassurance [19]. Model fact labels, which summarise intended use, training data and known limits, offer a more modest form of transparency [23].

D. Privacy and Security

Training and operating diagnostic AI requires access to sensitive records. Risks include unauthorised access, re-

identification of supposedly anonymised data, and misuse in sharing between institutions. In India, the Digital Personal Data Protection Act, 2023 sets consent-based rules for processing personal data [22]. The WHO lists protection of autonomy and privacy among its core principles for health AI [20].

E. Accountability and Integration

When an AI-supported diagnosis is wrong, liability could attach to the clinician, the hospital or the developer, and law has not settled the allocation. Practical barriers also slow adoption: integration with hospital information systems, staff training, cost, and uneven infrastructure in rural areas, which are precisely the places that could gain most from decision support [1], [20]. Models also drift as populations and equipment change, so performance must be monitored after deployment, not only at approval [24].

TABLE IV
LIMITATIONS AND COMMONLY PROPOSED SAFEGUARDS

Limitation	Typical safeguard
Biased or narrow data	Diverse datasets; subgroup reporting [16]
Weak external validation	Prospective multi-site trials [10], [14]
Opacity	Interpretable models; fact labels [18], [23]
Privacy exposure	Consent, access control, law [22]
Unclear liability	Defined clinician oversight [20]
Model drift	Post-deployment monitoring [24]

None of the safeguards is complete in isolation. Subgroup reporting exposes bias without removing it, trials are costly, and monitoring assumes that confirmed diagnoses are recorded. They are best treated as layers, each covering a weakness left by the others, in the manner of safety engineering in other high-risk fields.

F. Human Factors and Automation Bias

Technical accuracy does not settle how a tool behaves in a consulting room. Readers who see a prompt tend to agree with it, particularly when tired or when the case looks routine, so a wrong suggestion can pull a correct clinician toward error. The opposite failure also occurs: clinicians who learn that a tool raises many false alarms begin to ignore it, and a true alarm is then missed. Both effects are well known from alerting systems in general and argue for evaluating a tool together with its users, not apart from them [1], [4]. Design choices matter here. Showing the suggestion only after the clinician has formed an initial view, reporting calibrated confidence and keeping the alarm rate low are all measures a deploying hospital can control.

G. Responsibility for Updates

A model that learns from new data after release can improve, but it can also degrade without anyone noticing. Regulators have therefore begun to ask manufacturers to describe in advance how a device will be modified and monitored, so that changes follow a stated plan and not an ad hoc retraining [24]. For a hospital, this implies a standing duty to track the performance of every deployed tool on its own patients, by comparing the tool's outputs with eventual confirmed diagnoses and by watching for shifts in the case mix.

IX. GOVERNANCE AND RESPONSIBLE ADOPTION

Regulation is moving from principle to rule. The WHO guidance of 2021 sets six principles, including transparency, accountability and equity [20]. The EU Artificial Intelligence Act, which entered into force in August 2024, classes medical AI as high-risk and phases in requirements on data governance, documentation, human oversight and accuracy [21]. In the United States, the FDA clears AI devices through its existing medical-device pathways and has issued guidance on updates made after authorisation [12], [24]. India has data-protection law [22] but, as far as the sources reviewed here show, no dedicated statute for clinical AI.

Ethical principles become concrete only when attached to decisions. Autonomy requires that patients know when an algorithm contributed and can ask for a human opinion. Beneficence and non-maleficence require evidence that the tool improves outcomes and does not shift harm onto particular groups. Justice requires that benefits reach populations that lack specialists and not only well-resourced hospitals [20]. Transparency requires documentation that a clinician can read: intended use, training population, known failure cases and performance by subgroup [23]. These requirements overlap with the legal duties emerging in the European Union and elsewhere, and a hospital that meets the ethical ones will usually find the legal ones within reach [21].

A further question is the standard of comparison. A tool is often judged against the average clinician, yet the relevant comparison for a given hospital may be its own staff, or the alternative of no reading at all in a region without a specialist. The answer changes the ethical balance. A modest tool that replaces nothing and reaches patients who would otherwise receive no assessment may do more good than an excellent tool that duplicates an existing service [1], [20].

TABLE III
REGULATORY APPROACHES TO MEDICAL AI

Body	Instrument	Main emphasis
WHO	Guidance, 2021 [20]	Six ethical principles
European Union	AI Act, 2024 [21]	High-risk duties, oversight
United States	FDA device pathways [12]	Premarket review, updates
India	DPDP Act, 2023 [22]	Consent, data protection

The table shows that only the European instrument is written specifically for AI as a category; the others rely on guidance, on general device rules or on data law. A hospital operating across jurisdictions must therefore reconcile

several overlapping regimes, which favours organisations that already hold documentation of data sources, testing and monitoring.

Drawing on the synopsis and the literature, five practices recur: training on diverse, well-documented data; independent external and prospective validation; interpretable output with a stated intended use; strong privacy controls; and clinician oversight with defined accountability. None is novel, and their value lies in being applied together.

The Indian setting illustrates how these threads meet. The country has a large population, an uneven spread of specialists, and a growing digital health infrastructure, which together make decision support attractive. The same conditions raise the stakes of the limitations above. Models trained mainly on data from high-income countries may not transfer to local populations, equipment and disease patterns, and data collected at scale must be handled under the 2023 data-protection law [22]. Local validation before deployment is therefore a safeguard and not a formality [4], [20].

Education is the remaining piece. Clinicians need enough understanding to question an output: to ask what data the system was trained on, how it was tested, and where it is known to fail. Model fact labels and reporting checklists make that questioning practical [4], [23]. Patients, for their part, are entitled to know when an algorithm contributed to their diagnosis [20].

X. CONCLUSION AND FUTURE DIRECTIONS

AI-driven diagnosis has moved from research to regulated clinical products, with the strongest evidence in imaging, particularly screening for breast cancer and diabetic retinopathy. The documented benefits are faster reading, lower workload and consistent output. The documented limits are equally clear: biased data, thin external validation, opaque reasoning, privacy exposure and unsettled liability. The available evidence supports AI as a supervised aid and does not support autonomy beyond narrow, well-validated tasks.

Three longer-term questions deserve separate mention. The first concerns evidence of patient benefit. Most studies report diagnostic accuracy, which is an intermediate measure; fewer report whether patients live longer, suffer less or avoid unnecessary procedures. Over-detection is a known risk in screening, since finding more small abnormalities does not always mean helping the patient, and the same caution applies to algorithms that are very sensitive [10], [17]. The second concerns generalist models. Large multimodal systems promise a single tool across many tasks, but their reliability on any one task is harder to bound than that of a narrow classifier, and their errors can be fluent and convincing [13]. The third concerns sustainability. Maintaining a model, auditing it and retraining it costs staff time and money, and these costs fall hardest on small hospitals [5], [24].

Several directions follow from the gaps found. Multimodal models that combine images, records and text need prospective evaluation. Datasets from low- and middle-income settings are under-represented and should be built. Methods for interpretable and calibrated output deserve more attention than post-hoc explanation. Legal frameworks for liability need to catch up with practice. Because this paper is a review without original experiments, its conclusions are limited by the studies it cites, and the 2025 regulatory figures should be rechecked against the sources before reuse.

REFERENCES

[1] E. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York, NY, USA: Basic Books, 2019.
[2] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, pp. 24–29, 2019.
[3] P. Rajpurkar *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
[4] Y. Liu, P.-H. C. Chen, J. Krause, and L. Peng, "How to read articles that use machine learning: Users' guides to the medical literature," *JAMA*, vol. 322, no. 18, pp. 1806–1816, 2019.
[5] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nat. Biomed. Eng.*, vol. 2, pp. 719–731, 2018.
[6] H. Singh, A. N. D. Meyer, and E. J. Thomas, "The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations," *BMJ Qual. Saf.*, vol. 23, no. 9, pp. 727–731, 2014.
[7] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
[8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
[9] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, pp. 89–94, 2020.
[10] K. Lång *et al.*, "Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): A clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study," *Lancet Oncol.*, vol. 24, no. 8, pp. 936–944, 2023.
[11] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," *npj Digit. Med.*, vol. 1, art. 39, 2018.
[12] U.S. Food and Drug Administration, "Artificial intelligence-enabled medical devices," FDA, Silver Spring, MD, USA. [Online]. Available: https://www.fda.gov (accessed 2025).
[13] K. Singhal *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, pp. 172–180, 2023.
[14] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, 2023.
[15] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, art. m1328, 2020.
[16] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.
[17] A. Adamson and H. G. Welch, "Machine learning and the cancer-diagnosis problem — no gold standard," *N. Engl. J. Med.*, vol. 381, pp. 2285–2287, 2019.
[18] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, pp. 206–215, 2019.
[19] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *Lancet Digit. Health*, vol. 3, no. 11, pp. e745–e750, 2021.
[20] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva, Switzerland: WHO, 2021.
[21] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), *Off. J. Eur. Union*, Jul. 2024.
[22] Government of India, "The Digital Personal Data Protection Act, 2023," Act No. 22 of 2023.
[23] M. Sendak *et al.*, "Presenting machine learning model information to clinical end users with model facts labels," *npj Digit. Med.*, vol. 3, art. 41, 2020.
[24] R. Parikh, Z. Obermeyer, and A. Navathe, "Regulation of predictive analytics in medicine," *Science*, vol. 363, no. 6429, pp. 810–812, 2019.