SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper
Open Access

AI For Medical Diagnosis: Advances, Applications, And Future Directions

S. Sathyanarayanan[1], Abhijit Musale[2*], Suhas K C[3], Swathi P Sindhe[4], Shilpa Sthavarikmath[5]

[1,3] Sri Sathya Sai University for Human Excellence, Kalaburagi, Karnataka, India.
[2]Sri Sathya Sai Hospital, Belgaum District, Karnataka, India
[4,5] Central University of Karnataka, Kalaburagi, Karnataka, India
*Corresponding author: abhijitmusale.sssh@gmail.com

Abstract
AI (Artificial Intelligence) is significantly advancing clinical diagnostic processes. Recent progress in the use of AI for medical diagnostics, such as computer vision based on medical imaging and understanding physiological signs and symptoms (acoustic cardiac signals), is discussed in this review paper. A literature review was conducted covering the period 2016-2025, and the results of this research are presented, examining the achievements of AI systems, the issues encountered in the diagnostic process, and their ethical implications. From the results presented thus far, AI matches specialists on narrow, well-defined imaging tasks; approaches non-expert level for open-ended generative diagnosis; and still trails experts overall. One crucial and undeniable task of explainable AI (XAI) is to build trust in these systems among clinicians. Finally, some promising pilots of clinical engagement and research priorities are presented, along with the possibilities (and restrictions) of the capabilities that AI can bring to this field.

Keywords: Artificial Intelligence, Medical Diagnostics, Medical Imaging, Explainable AI (XAI), Clinical Decision Support.

1 Introduction

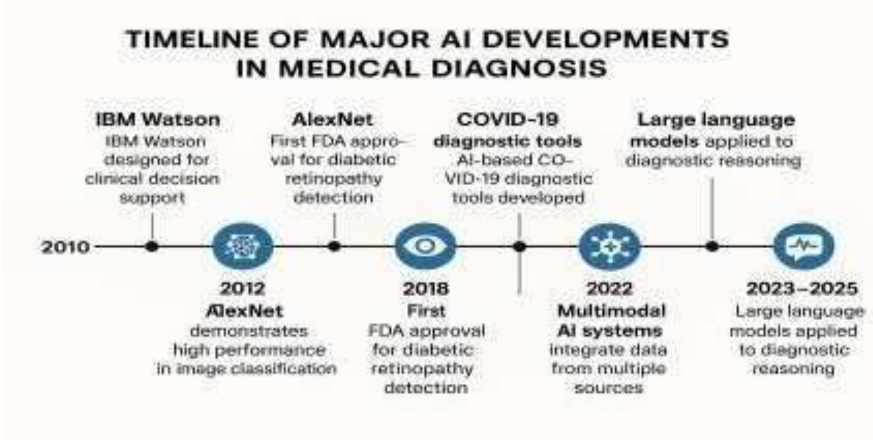
The growing demand for diagnostics, staffing shortages, and the increasing volume of medical information are causing strain on healthcare systems across the globe. A conservative estimate in the USA suggests that one in ten people experiences a significant diagnostic error each year, with errors resulting in 40,000 to 80,000 deaths per year [1]. These challenges have positioned AI as a leading solution for accurate and effective diagnosis, helping to bridge the healthcare divide. AI is already present in numerous fascinating areas of the medical sector, such as radiology, pathology, and electrocardiography (ECG), and AI systems allow for the analysis and interpretation of patient records. This paper aims to introduce the recent progressions in the domain of medical applications in diagnostics using Artificial Intelligence (AI). After establishing foundational concepts, evaluating AI performance in specific scenarios, and discussing challenges to clinical implementation, we introduce a framework and the resulting opportunities to integrate clinical AI, specifically for prioritisations. Guidelines are given to healthcare institutions, healthcare professionals and regulators based on the study.

2 Literature Review

2.1 Historical Development of AI in Medical Diagnosis

AI has been around since the 1970s, including MYCIN, one of the first AI-based medical diagnosis programs for infectious disease diagnosis (IDD). The drawback of these rule-based systems is that only hard-coded knowledge is stored, and they lack the ability to learn. In the 1990s, new algorithms for pattern recognition (using machine learning (ML)) were developed for diagnostics, such as Bayesian networks and support vector machines (SVMs), for more efficient pattern recognition. However, this shifted in the 2010s, when deep learning (DL), particularly CNNs, were introduced for image processing. In 2012, at the ImageNet competition, AlexNet achieved much better performance on image classification problems. Its use was immediately initiated in medicine and was soon seen to have positive results in dermatology, ophthalmology and radiology. The success of DL algorithms has been demonstrated to be comparable to that of experts in various applications, such as the identification of DR in retinal images since 2016 [2] [3]. The Medical Diagnosis Milestone chart (2010-2025) shown in Figure 1 illustrates the timeline of major developments in the field of medical diagnosis using Artificial Intelligence between 2010 and 2025.

Figure 1. Timeline of Major AI Developments in Medical Diagnosis (2010-2025)



2.2 Recent Research Trends (2023-2025)

Enormous progress was made in AI diagnostics between 2023 and 2025. A system that integrates multiple types of data sources—such as imaging, clinical notes, and genomics—to perform diagnostic tasks is considered a multimodal system. First, multimodal systems developed to fulfil more than one diagnostic task have replaced previous narrow systems, which are mostly based on images. Second, the difficulty of explainable AI (XAI) was significant, and researchers and clinicians wanted to understand how AI systems get to their conclusions from the data. There are also some applications of interpretable models in the healthcare sector, such as the work done by Chanda et al [4]. The performance of the AI model with XAI was also demonstrated to be better in a clinical study, with an improvement of 2.8% in the test of balanced accuracy for the diagnostic task compared with that of the conventional AI model. Third, large language models (LLMs) have made an impact in diagnostics. However, the benefits of these methods are not guaranteed. In a randomised trial, Goh et al. concluded that there was no significant difference in the diagnostic performance of physicians who were given access to ChatGPT [5]. A fourth method, Federated Learning, provides collaborative models to multiple organizations without sharing raw patient data, thereby maintaining privacy. This has allowed for the creation of more generalisable models, called "models based on heterogeneous populations”.

2.3 Research Gaps and Opportunities

However, there are still several fields of study that need to be addressed. One of these is that none of the AI diagnostic systems in the literature can generalise because most of them are trained and tested in high-resource environments with a homogeneous set of people. Second, few studies have assessed the efficiency of AI diagnostic instruments in clinical outcomes in longitudinal studies. Third, all existing evaluation systems are quite specific and do not apply to all types of AI and expert systems. These gaps can be used as a starting point for future studies, such as creating AI systems that are specific to resource-constrained environments and exploring the best models of human-AI collaboration.

3 AI Technologies in Medical Diagnosis

3.1 Machine Learning Algorithms and Their Diagnostic Applications

Random forests have proven to be easy to use for predicting diagnoses. Different implementations of ML models for the prediction of heart disease were analysed, and it was observed that ensemble methods may be able to provide the highest percentage of accuracy of the prediction of the disease, that is, 92%, as reported by [6]. Unsupervised techniques have been effective in discovering and even recognising new sets of patterns and subsets. They tend to be most helpful for patients with complex conditions (such as autoimmune diseases) when conventional tests fail to detect the underlying cause of their illnesses. Although currently rarely applied, optimizing diagnostic workflows and decision-making processes using reinforcement learning shows significant promise.

3.2 Deep Learning Architectures and Multimodal Approach for Medical Image Analysis

Various studies have used multimodal data for diagnoses. For example, Poterucha et al. created a DL model called EchoNext, which can be used to detect cardiovascular diseases from over one million ECG data points. The DL model outperformed cardiologists in diagnosing the disease using ECG data. Multimodal systems can be used for various diseases with numerous symptoms, such as autoimmune diseases, cancers, and neurodegenerative diseases. This is because multimodal systems use multiple source data systems, which allow the discovery of smaller correlations or relations that could not be discovered in the case of the use of a single mode.

3.3 Sequential and Audio Processing Architectures

Diagnostic methods requiring continuous physiological signals, such as electrocardiograms (ECG) and phonocardiograms (PCG), require models which can model temporal dependency. Time-series medical data were processed using Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. In recent years, 1D-CNNs and Temporal Convolutional Networks (TCNs) have been shown to be competitive on the continuous scale for medical signal processing in the field of electrical and acoustic signals. Furthermore, by customising the audio transformer, advanced extraction of time-related features in digital auscultation has become possible, and AI systems can detect subtle physiological abnormalities that are likely to change over time.

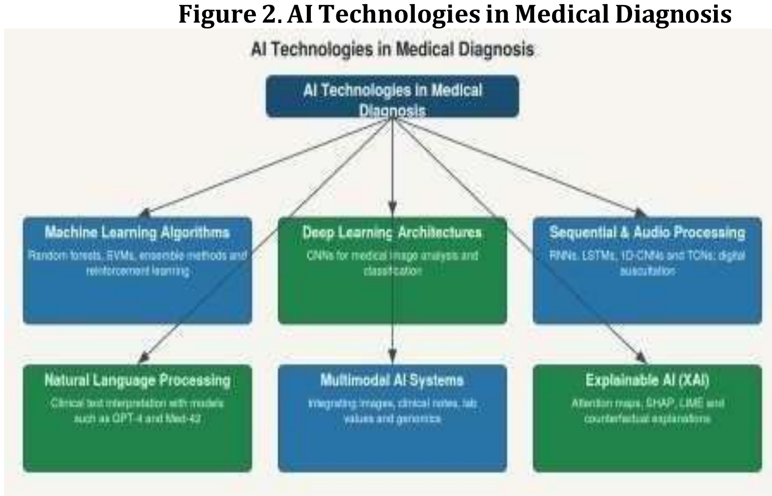
3.4 Natural Language Processing for Clinical Text Interpretation

Natural Language Processing (NLP) has made significant recent advancements in diagnosing from various clinical texts, including radiology reports, pathology reports, and clinical narratives. There are also models, such as Med-42, that can perform medical-related tasks, and models such as GPT-4 that can read and reason with medical jargon and medical diagnostic issues. Recently, Takita et al. compared the diagnostic capacity of various LLMs with non-expert and expert physicians, and in some diagnostic scenarios, no difference was found between the diagnostic accuracy of LLMs (such as GPT-4o, Llama3 70B, and Claude 3 Sonnet) and non-expert physicians, while LLM performance was worse than that of expert physicians ($p = 0.007$) in others [7]. Another use of NLP is in computerising the process of automatically extracting diagnostic criteria from clinical practice guidelines and medical literature to form continuously updated diagnostic decision support systems. This property will be useful in areas where the diagnosis may change as evidence accumulates quickly, such as in the field of oncology.

3.5 Explainable AI Approaches for Clinical Decision Support

In the area of XAI, attention is paid to the black box problem and providing interpretation to the outputs. In the diagnostic context, this has been accomplished using attention maps, feature importance scores, and the provision of counterfactual explanations. Chanda et al. showed that explainability increased dermatologists' confidence in the decisions and diagnostic accuracy of the dermatology system [4]. SHapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are popular methods for obtaining explanations for individual predictions and identifying the features that play a major role in the prediction [4].

Figure 2. AI Technologies in Medical Diagnosis



4 Specialty-Specific Applications

4.1 Radiology

Radiology is one of the most established fields of diagnostic AI application [8]. The example of the implementation of AI in screening mammography is a study conducted by McKinney et al., who developed a system that achieved fewer false positives and false negatives compared to radiologists in large cohorts of the UK and US populations [9]. The following MASAI trial, which was the first randomized controlled trial on the use of AI in mammographic screening and included more than 100,000 women, revealed that the use of AI-assisted reading-maintained cancer detection rates and lowered radiologists' screen-reading burden by 44% [10]. In case of lung cancer, Ardila et al. developed an end-to-end three-dimensional model that can estimate the risk of malignancy based on low-dose chest computed tomography and showed 94.4% area under the curve and equal or better performance compared to six radiologists [11].

4.2 Pathology

AI-based analysis has revolutionised histopathological diagnosis. AI has been proven to be useful for cancer detection, grading, and the quantification of cancer biomarkers. In breast cancer, Challa et al. validated a commercial AI tool (the Visiopharm Integrator System) for detecting lymph-node metastases on haematoxylin-and-eosin whole-slide images within a clinical digital workflow: the algorithm identified every metastasis (100% sensitivity and negative predictive value), and pathologists reviewing its annotations reached 99% concordance with the reference diagnosis [12]. An example of human-centred AI in the field of pathology is the design of Nuclei.io by Huang et al. from Stanford Medicine [13]. It is flexible, gains a dedicated workflow, and provides customised help to find cells that are of concern for the disease. It can also contribute to teamwork; pathologists can download models via web technology and learn from their colleagues by interacting with each other. The two primary challenges in integrating AI into the pathology workflow are the resolution and size of the whole slides to be processed and the complexity of the diagnostic task. The integration was done using a hierarchical approach; low-magnification images were used to identify the regions of interest, and detailed views were taken at high magnifications to diagnose the region of interest. It reduces the number of calculations while providing a fairly accurate diagnosis.

4.3 Dermatology

AI systems can perform skin lesion classification at the level of physicians. Chanda et al. performed an eye-tracking study in which 76 dermatologists diagnosed dermoscopic images with the aid of an XAI system [4]. A 2024 study on the classification of skin lesions by Salinas et al. determined a clinically similar performance when comparing AI-enabled systems and expert dermatologists, implying that AI technologies can be a valuable add-on to dermatological practice, particularly in the underserved regions [14]. Mobile applications have democratised dermatological screening, enabling effective telemedicine. Marri et al. reported the accuracy of dermatological diagnosis using the Aysa app to be an average of 96.1% in the top three diagnoses and 80.6% for exact diagnosis [15].

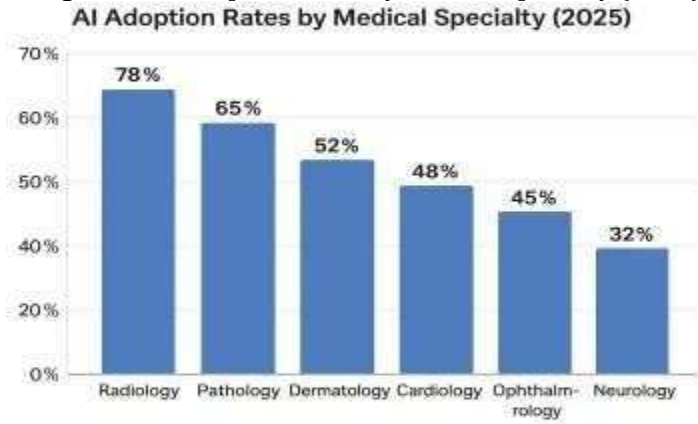
4.4 Cardiology

Imaging and non-imaging data are analysed by systems with the help of AI. DL models have been proven to be superior to nonspecialists in ECG interpretation. In a recent article, Wu et al. reviewed several real-world applications of DL for cardiovascular diseases in various scenarios [16]. Revuri et al. proved that among all AI-ECG models, it had the overall highest accuracy of 85.7% while assessing cardiac dysfunction from LVEF (left ventricular ejection fraction) [17]. In the same vein, another AI model for stress ECG interpretation for both male and female patients resulted in a substantial improvement in diagnostic accuracy from 50% to 90.9% [18]. Another such field is cardiac diagnostics, which has recently experienced a resurgence in the use of AI for analysing heart sounds via PCGs and digital stethoscopes. DL models are found to be very effective in the detection of an abnormal sound/murmur from 1-dimensional audio to 2-dimensional spectrograms [19], [20], [21], [22]. These non-intrusive, portable, and cost-effective acoustic applications are ideal for screening purposes. Standardised reporting and structured data collection help with the integration of AI into cardiac imaging workflow. Where dynamic imaging is crucial, the early experience of the use of 4D imaging and modelling of time is encouraging.

4.5 Other Specialties

To understand the importance of the DL model in the diagnosis of neurological diseases, Prasad et al. proposed an Alzheimer's disease classification system using only image data [23]. AI systems can identify DR and AMD and assist in the screening of glaucoma. AI has also boosted the ability of gastroenterologists to identify polyps and early cancer. During colonoscopy, the technology detects polyps with a sensitivity greater than 90% to reduce the risk of missing polyps. Other uses are still to come, such as analysing speech and facial expressions for psychiatric diagnosis, automatically scoring sleep studies for sleep disorders, and computer-assisted scoring of skin biopsies for rare skin diseases. The rate of adoption of AI by medical specialty is shown in Figure 3 below.

Figure 3. AI Adoption Rates by Medical Specialty (2025)



Areas that are more focused on direct patient care are also using Artificial Intelligence early on. There are factors that affect how quickly Artificial Intelligence is adopted, and these factors differ across states. For example, the difficulty of a diagnosis, problems with rules and laws, and whether we have data are all important factors.

5 Performance Analysis and Evaluation

5.1 Comprehensive Meta-Analysis of AI Diagnostic Performance

An extensive study was conducted on all available literature, and a meta-analysis was performed by Takita et al. in 2025 to compare the efficacy of Generative AI against physicians [7]. Overall, the Generative AI models achieved an accuracy of 52.1%. The actual accuracy ranged from 47% to 57.1% (95% confidence interval). We examined 83 studies conducted between June 2018 and June 2024 to obtain these numbers for the Comprehensive Meta-Analysis of AI Performance. Meta-analyses indicate that while AI models demonstrate diagnostic accuracy comparable to non-expert physicians ($p = 0.93$), expert physicians still perform significantly better, achieving 15.8% higher accuracy than AI systems [7]. Salinas et al. studied the classification of skin lesions and found that AI algorithms are as accurate as expert dermatologists [14]. AI systems with sensitivities and specificities comparable to radiologists' accuracy were found to be quite accurate for common diseases but not for uncommon diseases, as pointed out by Khalid et al. [24]

5.2 Comparison Between AI Systems and Healthcare Professionals

There were minor discrepancies between AI systems and healthcare providers in other specialties across the board in terms of their knowledge and experience. According to Goh et al., in a randomised trial (known as the Stanford HAI study), GPT-4 achieved a score of approximately 80% on the diagnostic reasoning task, whereas physicians using traditional resources obtained scores of approximately 70% [5]. Perhaps most significantly, the models were not significantly better than the physicians without access to the model, indicating that one of the greatest difficulties was human–AI interaction rather than AI capability.

5.3 Factors Affecting AI Performance

The experts pointed out some subtleties regarding the performance of various specialties and expertise levels (whether those of AI or of a healthcare professional). AI is more effective than clinicians without AI assistance for repetitive and standardised tasks, can complement and augment clinical judgment in complex and rare cases, and is most useful in conjunction with clinical judgment.

Figure 4. AI Diagnostic Performance

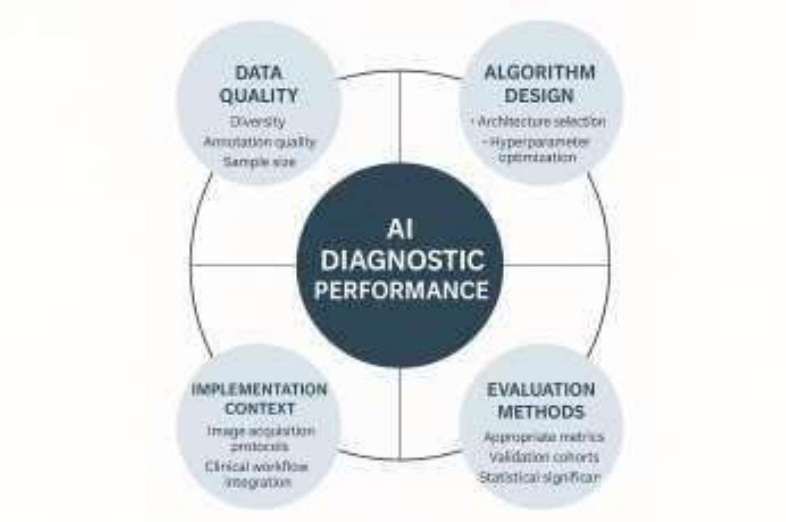


Figure 4 illustrates the AI diagnostic performance. Several factors can affect the performance of an artificial intelligence model. El-Khoury and Zaatari investigated the diagnostic accuracy of models across different patient subgroups by race, insurance type, and age [25]. The results were quite diverse across these patient subgroups, emphasising the need for training models with diverse patient data. The architecture is of great significance. While transformer-based models are effective for multimodality integration, specialised architectures such as U-Net are more effective for segmentation, and ensembles are more effective for complex diagnostic tasks. The context of deployment is also critical; the image acquisition protocols used, image processing, and model integration into clinical workflows all impact the results. The clinical information-based model was superior to the imaging-based model. As shown in Figure 4, various multidimensional factors must be considered in the clinical setting when assessing the use of AI technology for diagnostics. The largest domain is the AI diagnostic performance, which is affected by size, quality and diversity of data, model architecture and the deployment context. The technical difficulty of visual imaging is complicated by the need for continuous monitoring of human physiology. Audio data is sensitive to factors that affect the signal-to-noise ratio, meaning that if the AI was trained on 'clean' audio without any pre-processing steps during the audio recording process, it will not be able to transfer its knowledge to 'dirty' clinical recordings.

5.4 Standardised Evaluation Metrics and Benchmarks

The evaluation method has now been standardised. The most common measures are sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Various measures, including precision-recall curves, F1 scores, and calibration measures, have been reported and are currently being used. Some benchmark datasets have been published, including MIMIC-CXR for chest radiographs, HAM10000 for skin lesions, and ADNI for neuroimaging data from Alzheimer's disease. However, most are deficient in the diversity of patient cases and settings, limiting the generalisations that can be made from the results. Today, the focus has been on developing more complex types of evaluation systems whose goals include being diagnostic, scale-free, demographically representative, and capable of accommodating the calibration of uncertainty estimates.

5.5 Statistical Analysis of Diagnostic Accuracy Across Specialties

The follow-up statistical research results showed that across the board there was a statistically significant change between the traditional clinical diagnosis and the post-AI intervention. Moreover, in diagnosis-heavy specialties (such as radiology, ophthalmology, and dermatology), AI systems perform comparably or favourably against traditional diagnostic methods. Takita et al. (2025) explained in their analysis that AI diagnostic performance did not differ significantly across most specialties, with dermatology and urology as exceptions. The effectiveness of AI may vary depending on the field of expertise, type of AI, the ways in which visual pattern matching can aid the diagnosis process, and the level of standardisation of the data. There is good news for those with imaging and visual signs (e.g. ophthalmology and dermatology) and not so good for those based on many clinical factors (e.g.psychiatry and rheumatology). The overall trend of accuracy has been generally increasing over the years, as the accuracy by specialty has improved by an average of 3-5% per year between 2020 and 2025. The improvement is the result of improvements in the algorithms, availability of training data, system changes, and improvements in system evaluation.

6 Clinical Integration and Implementation

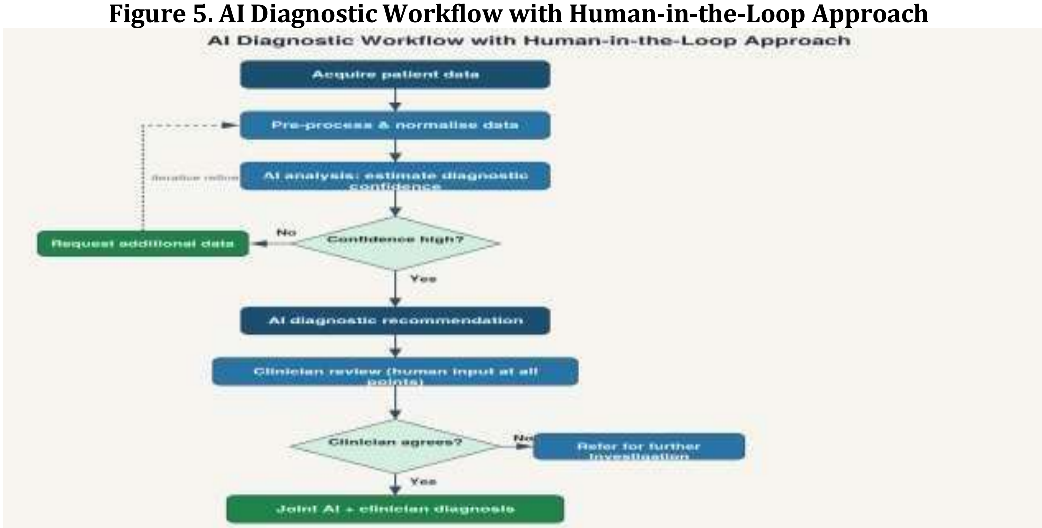
6.1 Regulatory Frameworks and Approval Processes

Regulatory guidance has emerged to incorporate information from AI diagnostic systems. The FDA's Digital Health Software Pre-Cert Pilot (2017–2022) conducted trials assessing developer quality and organisational excellence rather than reviewing each device individually, but the pilot was closed in September 2022 after the FDA determined it could not be implemented under current statutory authority. Most AI diagnostic products are either Class II or Class I, meeting the moderate-risk criteria of the European Union Medical Device Regulation (MDR). Obuchowicz et al. observed that most radiology AI products fall within these moderate-risk MDR classes — a classification that balances regulatory oversight with innovation while safeguarding patient safety [8]. The International Medical Device Regulators Forum (IMDRF) is developing provisions for medical devices with AI to support their growth and minimise regulatory barriers, thereby developing a common and clear regulatory framework for their introduction worldwide [26].

6.2 Human-AI Collaboration Models and Workflow Integration

The managerial implementation of the model will guarantee the creation of tools that will work well with both humans and AI systems. The human-in-the-loop method will make it more probable that an AI will be viewed as a "clinical assistant", as in the example of the Stanford pathology tool Nuclei.io [13]. This approach considers the freedom of medical professionals, the benefits of utilising the potential of AI, and the accuracy and effectiveness of the system. AI systems have been successfully used in radiology for parallel reading workflows: the images are forwarded to the AI system, and a suspected diagnosis is forwarded, and the AI system highlights the suspected diagnosis. Through this type of integration, a Northwestern Medicine AI method has led to a 15.5% increase in the efficiency of radiographer reporting. The AI triage model was found to be useful for bulk screenings.

Figure 5. AI Diagnostic Workflow with Human-in-the-Loop Approach

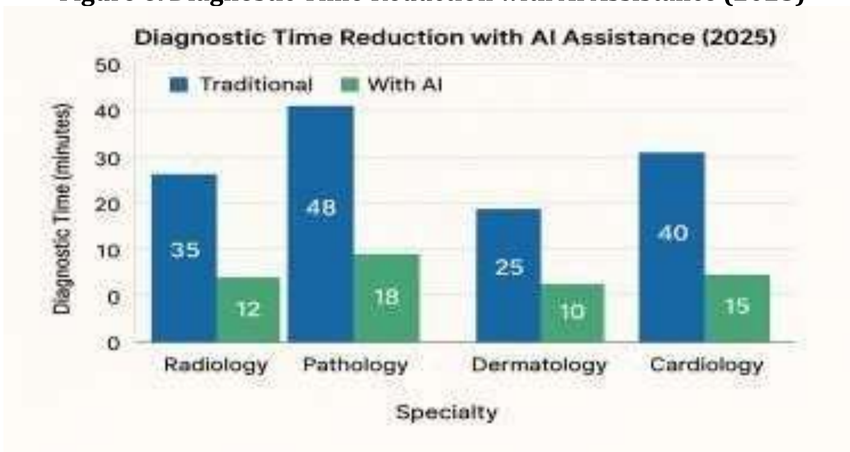


The model selected for implementation in a clinical setting was the human-in-the-loop AI diagnostic workflow shown in Figure 5. The steps include acquiring patient data, which are then pre-processed to normalise the data for analysis by AI. The diagnostic confidence of these data was then input into an AI system. The system offers diagnostic recommendations for obtaining good results. If the result is not highly confident, further enquiries for data are requested, and there is an iterative improvement process. Even better, the AI recommendations are brought to the attention of clinicians, and human input is always required. Finally, if the clinician(s) agrees with the AI diagnosis, the clinician(s) and AI provide the diagnosis together. If not, the case is referred for further investigation and may lead to a new diagnosis. This collaborative approach preserves the clinician's independent prescribing authority while utilizing AI computational engines responsibly, thereby mitigating the risks of automation bias. The method has been subjected to studies which have shown improvements in terms of accuracy and high satisfaction from clinicians.

6.3 Training Requirements for Healthcare Professionals

AI is becoming a universal language in the medical world, not only by understanding the concepts and employing them with critical consideration but also by applying AI in healthcare environments. Many of these medical AI ideas have gradually been integrated into an AI Literacy course designed to build awareness and thoughtfulness of AI-generated material and to use AI tools in the healthcare setting. Continuing medical education is increasingly focused not only on the technical underpinnings of AI, but also on its practical implementation across various medical specialties. Having the ability to do simulation training has been helpful in anticipating where it might go wrong and how to deal with it. There is a diagnostic time reduction with the use of AI.

Figure 6. Diagnostic Time Reduction with AI Assistance (2025)



It is evident from Figure 6 that the diagnostic time required across these four specialties significantly decreased when using AI assistance. This can reduce diagnostic delays and ease specialist workload [27]

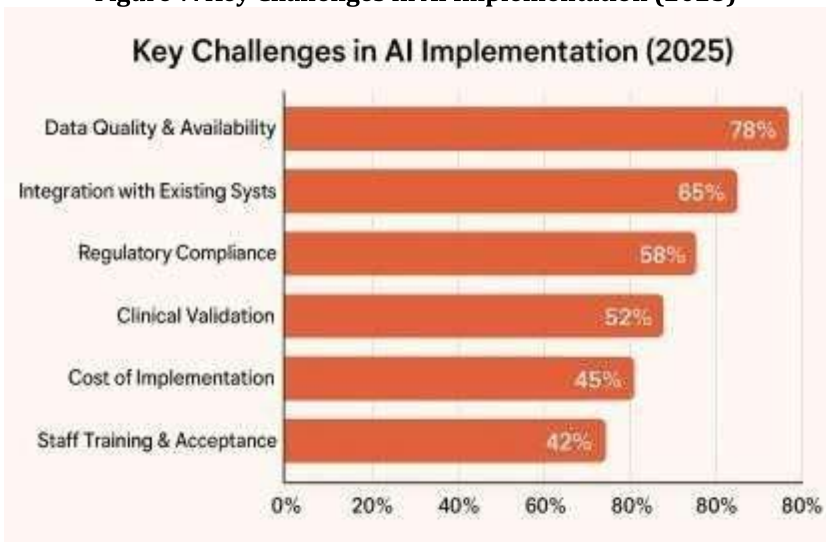
6.4 Case Studies of Successful Clinical Implementations

Several case studies have reported successful clinical outcomes. A 60% drop in the time required to diagnose a stroke was essential for the success of an AI system implemented at the Mayo Clinic to detect stroke on CT, including the results from the active involvement of stakeholders and a comprehensive redesign of the stroke workflow [28]. Intermountain Healthcare's addition of the AI retinopathy screening tool to their primary care platforms is another great example of where specialist-level care can be extended to populations that need it, as the company was able to augment the screening process and boost screening rates while maintaining the diagnostic accuracy of specialist-level screening [29]. The AI-enabled prostate cancer reporting workflow at Mount Sinai Health System reduced the prostate cancer reporting time by 40 percent while enhancing the consistency of Gleason scoring [30]. This success was achieved by engaging many pathologists in the design of the system and making improvements based on their feedback.

6.5 Lessons Learned from Failed Implementations

The failure cases studied provide valuable information for future studies. Challenges included a lack of a perfect fit between AI recommendations and the implementing hospitals due to cultural and geographical variations. For example, IBM Watson Oncology failed to produce the desired results because there was no real-world validation of its treatment plans across diverse patient groups. Failure due to the lack of workflow integration, not engaging the right stakeholders, or unrealistic expectations of what AI can do are other reasons for this. The dangers are all too familiar: technology first, people and organisations second; neglecting data quality and data standardisation; and neglecting maintenance and updates. To achieve effective implementation, it is important to have end-users involved from the beginning, with steps to move forward and apply them in an incremental way, set up good monitoring systems, and conduct targeted training [31].

Figure 7. Key Challenges in AI Implementation (2025)



The most important barriers that healthcare institutions encountered in the adoption of AI technology were issues of data quality/availability, integration of AI with IT infrastructure, and regulatory challenges, followed by clinical validation requirements, costs, and staff training, as depicted in Figure 7. The architecture of the clinical integration of AI diagnostic systems is illustrated in Figure 8.

Figure 8. Framework for Clinical Integration of AI Diagnostic Systems

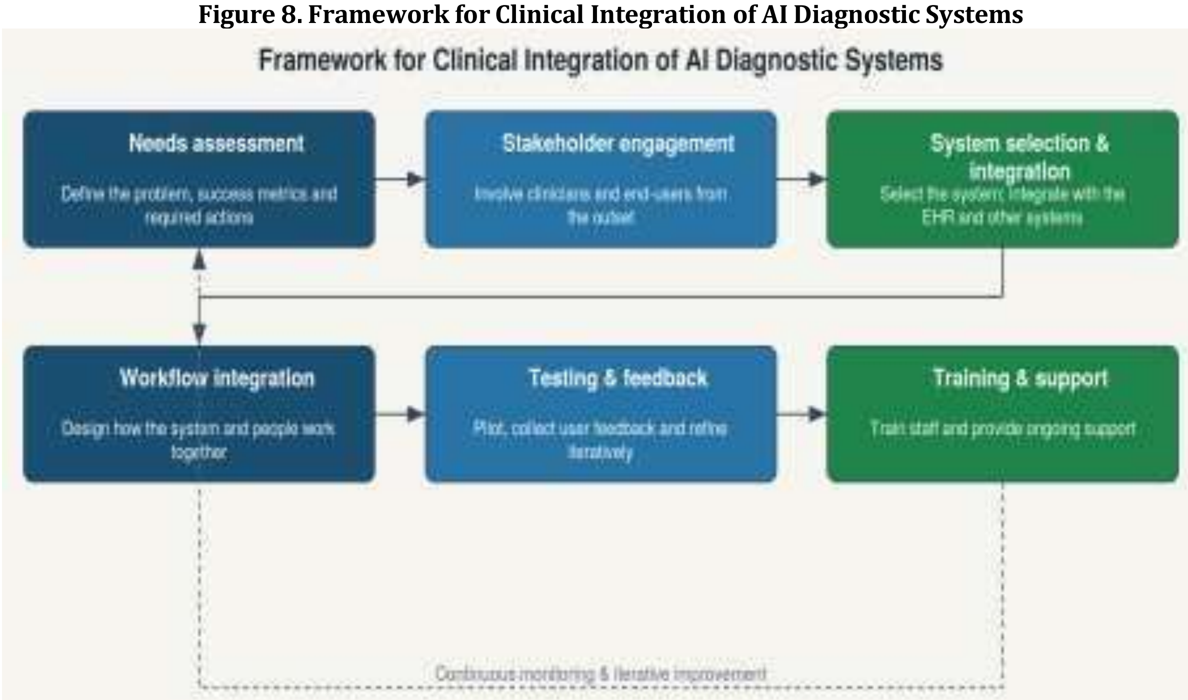


Figure 8 presents a six-stage framework for integrating AI diagnostic systems into clinical practice. The first stage is a needs assessment that defines the problem to be addressed, the outcome measures to be used, and the goals to be achieved. The second is stakeholder engagement: consulting end-users and incorporating their feedback early is essential for later adoption. The third stage is system selection and technical integration, ensuring interoperability with the electronic health record (EHR) and other clinical systems. The fourth — and most challenging — stage concerns the human–AI interface: how clinicians interact with the system in practice, refined through testing and iterative user feedback. The fifth stage provides training and ongoing support so that staff can use the tool competently. Finally, sustained dialogue among clinicians, technicians, and administrators treats implementation as a shared, institution-wide effort rather than a purely technical deployment.

7 Ethical Considerations and Challenges

7.1 Algorithmic Bias and Health Disparities

Bias can occur in several ways: imbalance in some groups in the dataset used for training the model; distortions in the model dataset from underlying concepts that are carried over to clinical practice; and use of proxy variables for protected characteristics in the model. This may result in delayed and/or undiagnosed identification of marginalised groups. Some suggestions for reducing bias are having a diverse training set, algorithmic fairness (e.g. adversarial debiasing), fairness constraint in model training, reviewing the performance of the model periodically by demographic groups, and reporting the model performance by demographic groups.

7.2 Privacy and Data Security Concerns

The use of AI diagnosis systems poses a variety of obvious privacy and security problems, including unauthorised access, re-identification of the underlying disguised client data, and misuse. Federated learning algorithms have proven to be effective because they can be applied to train models in different institutions. This does not reveal any information or improve the quality of the model. However, by carefully designing the noise, differential privacy provides mathematical guarantees that its training performance is not affected, while it is difficult to re-identify the individual. It has been discovered to work effectively in various diagnostic systems, such as deciphering X-rays during chest examinations.

7.3 Liability and Responsibility in AI-Assisted Diagnosis

The issue of liability and responsibility of AI diagnosis systems is complex and multifaceted. Who will be liable in case of a diagnostic error in the use of AI systems: the developer, the healthcare institution, the clinician, or none of them? Is it reasonable to consider AI disclosures as part of the informed consent in the diagnosis process? What constitutes the standard of care when utilising AI tools? The law is evolving and differs from state to state. Some systems rely on the 'learned-intermediary doctrine,' where the healthcare provider has the final say, whereas others lean toward shared accountability depending on the AI's level of autonomy. Professional societies have begun to provide guidance on the use of AI diagnostic tools and pertinent standards of care.

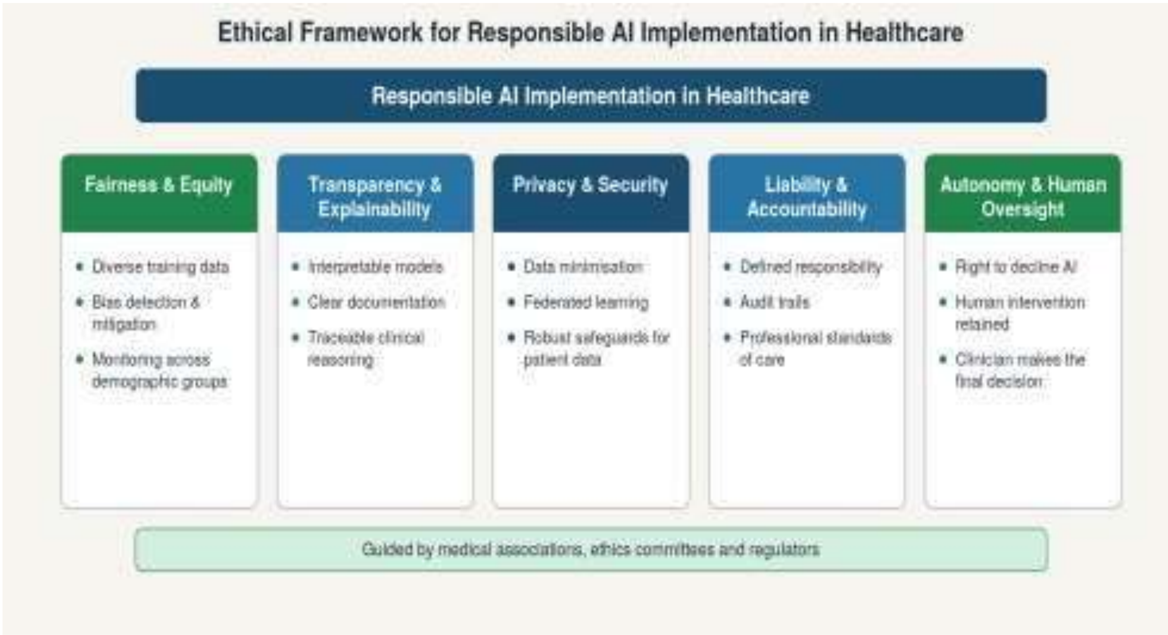
7.4 Transparency and Explainability Requirements

The principles of explainability and transparency are core to the use of AI in medicine. Most DL methods are black-box models and are thus not amenable to assessment in evidence-based practices and informed decision-making. Regulators are seeking an explanation. According to the AI Act, the EU will require technical documentation of AI in healthcare systems, which will classify some of these models as high-risk AI systems that must be assessed for safety and efficacy. In dermatology, the explainability has been shown in the enhanced accuracy and self-confidence of dermatologists by Chanda et al., and it was hypothesised to play a role in the interactions between humans and AI [4].

7.5 Patient and Provider Acceptance Factors

To be effective in combating errors in diagnosis, patients and providers need to adopt AI-powered diagnostic tools. Patients' general concerns include privacy, the possibility of being dehumanised, and concerns about the reliability of AI. Workflow, liability loss, and potential competition from professionals are providers' concerns. It is measured through a survey that is accepted or not based on certain factors such as perceived usefulness, ease of use, social influence, facilitating conditions, and trust. The importance of trust, especially, is at the centre of the focus, which is based on the openness of the system, its reputation, and its correlation with users' values. Focusing on what AI is capable of and what it is not, separating human tasks and AI tasks, illustrating with tangible examples how AI can help providers and/or patients, making the end user an advisor on development and use of the AI system, and continuous education of the AI system are some strategies that can be followed to ensure the acceptance of AI

Figure 9. Ethical Framework for Responsible AI Implementation in Healthcare



First, we investigated the moral problems of the proper use of AI healthcare, focusing on five principles (see Figure 9). Fairness and equity highlight the value of a variety of training data, bias detection and reduction plans, and ongoing performance measurement among demographic groups, as it is essential to guarantee that AI systems are equally beneficial to all patient groups. It is the clarity and explainability that highlights the importance of modelling it in an interpretable fashion and a guideline of clear documentation that allows clinical reasoning and trust in the AI suggestions. Privacy and security principles involve ensuring data minimisation practices, a federated learning framework in which sensitive data are not moved or stored away from their local place of origin, and a robust infrastructure to safeguard patient information. These encompass liability/accountability, audits, and professionalism in AI implementation in clinics. Disclaimer: Clinicians retain the right to override or decline AI recommendations, and human oversight is always available. The advice of medical associations, ethics committees, and regulators was considered. Patient trust and clinician uptake are enhanced in organisations where ethics are instituted.

8 Future Directions

8.1 Emerging Technologies and Methodologies

Several emerging technologies are poised to reshape the diagnostic landscape. Foundation models are a key driver: pretrained on large corpora, they can be adapted to specific diagnostic tasks with relatively little task-specific data. Edge AI and TinyML are advancing rapidly, using techniques such as knowledge distillation and pruning to compress models for deployment on low-power microchips — enabling point-of-care, real-time tools such as smart digital stethoscopes, where tight hardware–software integration supports more personalised and accessible care. Neuro-symbolic approaches, which combine symbolic reasoning with neural networks, offer a route to more interpretable systems by pairing predictions with human-readable explanations. Further ahead, quantum machine learning — already explored in genomics, proteomics, and drug discovery — may eventually extend the analytical capacity available for complex diagnostic data. Together, these directions point toward diagnostics that are faster, more explainable, and more widely accessible.

8.2 Generative AI Applications in Medical Diagnosis

Generative AI and large language models (LLMs) are becoming critical diagnostic tools due to their advanced contextual understanding. AI tools, such as generative AI, can assist with data compilation and input into models. It can even be set up to produce files for radiology and products for educating patients about their ailments. Some researchers have worked hard to delve into these images of tissues and create the Histopathology AI Explorer, developed by Ma et al. in 2025 [32]. Another possibility that may prove effective is the use of various types of generative techniques. In 2025, Hong et al. devised a model to make accurate prediction readings that could be directly applied in medicine [33]. These have the potential to be quite helpful in practice when using generative AI and LLMs.

8.3 Federated Learning for Privacy-Preserving Model Development

Federated learning: the ability to create AI models without exporting raw patient data from multiple institutions. Recent progress includes the design of personalised federated learning that considers the breadth of institutions in the number of treated patients and how they practice, secure aggregation methods that improve privacy guarantees, and approaches to work with non-IID (non-independent and identically distributed) data across multiple sites. In practice, this is being realised through multi-institutional consortia — for example, GenoMed4All, which applies federated learning across European reference centres to improve diagnosis of rare haematological diseases, alongside national initiatives such as HealthChain in France and Trustworthy Federated Data Analytics (TFDA) in Germany. This type of work is especially valuable in cases of rare diseases, for which there is insufficient data in a single repository to create a useful model [34], [35], [36].

8.4 Research Priorities and Opportunities

AI is also making an impact in medical education. Students and residents can test their diagnostic skills in realistic settings and receive feedback from AI mechanisms or even generate multiple scenarios, including situations that trainees have never experienced before. Adaptive learning systems that use AI can adapt training in the medical field, helping to determine what students do not understand and fine-tune the material accordingly.
The return on investment (ROI) of AI in Healthcare is illustrated in Figure 10.

Figure 10. Return on Investment for AI Implementation in Healthcare

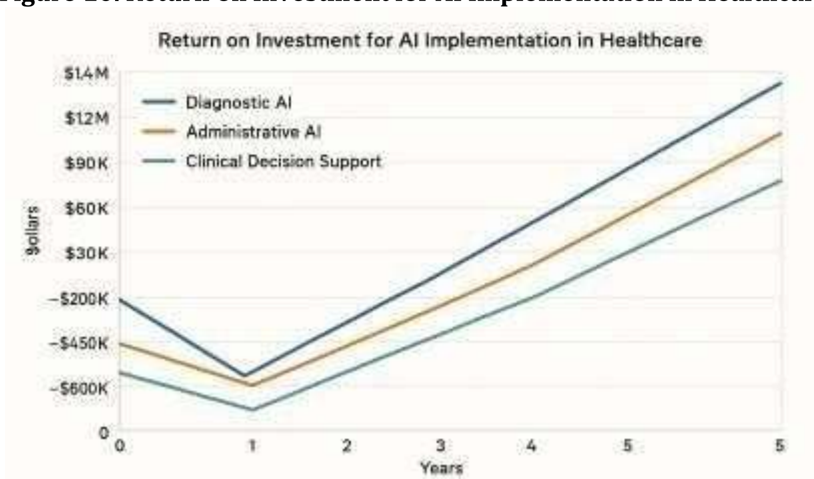


Figure 10 shows an example of ROI calculations for various implementations of AI across the diagnostic, administrative, and clinical decision support categories, based on their respective prices and expected lifetimes, indicating a relatively low initial cost for administrative implementations and the potential for a significantly higher ROI in the long term through the use of diagnostic implementations. Finally, there is a significant translational gap between commercially viable medical equipment and advanced AI prototypes developed in academic research settings. Algorithm research is blossoming; however, the theoretical development of clinical applications of an approved diagnostic tool is fraught with challenges, such as design patents for new AI-based diagnostic devices. Future research should be directed towards building strong university–industry partnerships, thereby speeding up the process. Research activity should be directed towards the priorities of the big scientific and medical research councils of the country (ICMR, DST, DBT, etc.) to make clinical trials and regulatory procedures a streamlined process, culminating in the introduction of these cutting-edge point-of-care devices into the medical system.

9 Conclusion

In many medical areas, AI has offered many benefits, particularly in the diagnosis of a vast array of diseases, and in certain cases, it is as effective as doctors. This review clarified the various challenges that must be considered during implementation, as well as their potential solutions. Although AI cannot replace human knowledge, it can greatly support it by being involved in an effective collaboration process using efficient collaboration models. The most effective use cases demonstrate that AI boosts some part of the human decision-making process to provide efficiency, consistency, and accuracy. First, ethical and practical issues regarding explainability exist in the development of clinical AI systems. Second, the performance of XAI for diagnostics is a challenge, emphasising the need for transparency in collaboration between AI and humans. Third, it is not only a cognitive bias but also an algorithmic one. Similarly, accessing the equity and generalisability of these diagnostic skills through the injection of representative data in algorithm construction and testing algorithms over sub-populations is a major focus. Finally, workflow, organisational issues, and stakeholder participation must be considered. This real-world validation is what separates successful diagnostic AI from cases like IBM Watson for Oncology, whose treatment recommendations at MD Anderson were never validated on real patients before the project was shelved. We recommend taking it slow and steady, conducting a needs assessment, consulting stakeholders first, conducting testing at smaller scales with the clinical population, and then moving on to larger-scale testing before implementation. This should be constantly monitored, and user feedback (modification and revision) should be included to ensure that it is functional and effective in the future.

Regulatory flexibility allows for innovation and patient safety, coordination across borders to enable global deployment, and adaptability to unique challenges. It will also help tackle the challenges of "continuously learning systems". The accuracy of diagnosis and efficiency of care will be vastly improved, and as AI continues to develop and grow, it will result in equity, transparency, and good human-AI interactions.

Declarations

Funding. The authors did not receive any specific funding for this study.
Competing interests. The authors declare no conflict of interest.
Ethics approval. Not applicable because this is a review of published material.
Consent to participate. Not applicable.
Data availability. Not applicable — none of the datasets were generated or analysed during this study.
Author contributions. All authors were involved in the conception, literature search, writing of the paper, review of the final draft of the paper and consent.

Table 1. Factors Affecting AI Diagnostic Performance

Factor Category	Specific Factors	Impact on Performance	Mitigation Strategies
Data Quality	Class imbalance	Reduced performance on rare conditions	Oversampling, data augmentation, weighted loss functions
	Annotation quality	Inconsistent ground truth affects learning	Multiple annotators, adjudication processes, quality control
	Dataset size	Insufficient data leads to overfitting	Transfer learning, data augmentation, federated learning
	Demographic representation	Biased performance across populations	Diverse data collection, fairness constraints, stratified evaluation
Algorithm Design	Architecture selection	Different tasks require specialised approaches	Task-specific architecture selection, ensemble methods
	Hyperparameter optimisation	Suboptimal settings reduce performance	Systematic tuning, automated optimisation
	Feature engineering	Inappropriate features limit learning	Domain-informed feature selection, representation learning
	Regularisation strategies	Overfitting or underfitting	Cross-validation, early stopping, dropout
Implementation	Image acquisition variability	Inconsistent inputs affect generalisation	Standardisation protocols, domain adaptation
	Integration with clinical workflow	User interface affects effective use	Human-centred design, workflow analysis
	Deployment environment	Hardware limitations affect performance	Model compression, edge computing solutions
	Monitoring and updating	Model drift reduces performance over time	Continuous evaluation, periodic retraining

Table 2. Key Implementation Challenges and Mitigation Strategies

Challenge	Impact	Effective Mitigation Strategies
Data Quality & Availability	High	Data governance frameworks, standardised acquisition protocols, synthetic data augmentation
Integration with Existing Systems	High	API-based integration, middleware solutions, phased implementation approach
Regulatory Compliance	Medium	Regulatory expertise involvement from project start, documentation automation, compliance-by-design
Clinical Validation	High	Multi-site validation studies, continuous performance monitoring, specialty-specific validation protocols
Cost of Implementation	Medium	Shared infrastructure models, cloud-based solutions, consortium approaches for smaller institutions
Staff Training & Acceptance	High	Simulation-based training, champion programmes, incentive alignment, continuous education

References

1. H. Singh, G. D. Schiff, M. L. Graber, I. Onakpoya, and M. J. Thompson, "The global burden of diagnostic errors in primary care," BMJ Qual. Saf., vol. 26, no. 6, pp. 484–494, Jun. 2017, doi: 10.1136/BMJQS-2016-005401.
2. 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, doi: 10.1001/jama.2016.17216.
3. S. Sathyanarayanan and S. Chitnis, "A Survey of Machine Learning in Healthcare," in Artificial Intelligence Applications for Health Care, I., M. K. Ahirwal, N. D. Londhe, and A. Kumar, Eds., Boca Raton: Taylor & Francis Ltd, 2022, ch. A Survey o, pp. 1–22. doi: https://doi.org/10.1201/9781003241409-1.
4. T. Chanda et al., "Dermatologist-like explainable AI enhances melanoma diagnosis accuracy: eye-tracking study," Nature Communications 2025 16:1, vol. 16, no. 1, pp. 4739-, May 2025, doi: 10.1038/s41467-025-59532-5.
5. E. Goh et al., "Influence of a Large Language Model on Diagnostic Reasoning: A Randomized Clinical Vignette Study," medRxiv, Mar. 2024, doi: 10.1101/2024.03.12.24303785.
6. W. Alsabhan and A. Alfadhly, "Effectiveness of machine learning models in diagnosis of heart disease: a comparative study," Scientific Reports 2025 15:1, vol. 15, no. 1, pp. 24568-, Jul. 2025, doi: 10.1038/s41598-025-09423-y.
7. H. Takita et al., "A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians," npj Digital Medicine 2025 8:1, vol. 8, no. 1, pp. 175-, Mar. 2025, doi: 10.1038/s41746-025-01543-z.
8. R. Obuchowicz, J. Lasek, M. Wodziński, A. Piórkowski, M. Strzelecki, and K. Nurzynska, "Artificial Intelligence-Empowered Radiology—Current Status and Critical Review," Diagnostics 2025, Vol. 15, Page 282, vol. 15, no. 3, p. 282, Jan. 2025, doi: 10.3390/DIAGNOSTICS15030282.
9. S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," Nature 2020 577:7788, vol. 577, no. 7788, pp. 89–94, Jan. 2020, doi: 10.1038/s41586-019-1799-6.
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, Aug. 2023, doi: 10.1016/S1470-2045(23)00298-X.
11. D. Ardila et al., "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nature Medicine 2019 25:6, vol. 25, no. 6, pp. 954–961, May 2019, doi: 10.1038/s41591-019-0447-x.
12. B. Challa et al., "Artificial Intelligence–Aided Diagnosis of Breast Cancer Lymph Node Metastasis on Histologic Slides in a Digital Workflow," Modern Pathology, vol. 36, no. 8, Aug. 2023, doi: 10.1016/j.modpat.2023.100216.
13. Z. Huang et al., "A pathologist–AI collaboration framework for enhancing diagnostic accuracies and efficiencies," Nature Biomedical Engineering 2024 9:4, vol. 9, no. 4, pp. 455–470, Jun. 2024, doi: 10.1038/s41551-024-01223-5.
14. M. P. Salinas et al., "A systematic review and meta-analysis of artificial intelligence versus clinicians for skin

cancer diagnosis," npj Digital Medicine 2024 7:1, vol. 7, no. 1, pp. 125-, May 2024, doi: 10.1038/s41746-024-01103-x.
15. S. S. Marri, W. Albadri, M. S. Hyder, A. B. Janagond, and A. C. Inamadar, "Efficacy of an Artificial Intelligence App (Aysa) in Dermatological Diagnosis: Cross-Sectional Analysis," JMIR Dermatol 2024;7:e48811 https://derma.jmir.org/2024/1/e48811, vol. 7, no. 1, p. e48811, Jul. 2024, doi: 10.2196/48811.
16. Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," BioMedical Engineering OnLine 2025 24:1, vol. 24, no. 1, pp. 23-, Feb. 2025, doi: 10.1186/S12938-025-01349-W.
17. N. Revuri, Q. Dai La, M. Faltas, F. Pryor, S. Aradhya, and J. S. Kahlam, "Investigating the Efficacy of AI-Powered Innovations in ECG Analysis and Continuous Heart Monitoring: A Comprehensive Narrative Review," 2025, doi: 10.7759/cureus.89743.
18. A. Banerjee et al., "Enhancement of Stress ECG Performance with Machine Learning: A Single-Center Study," JACC: Advances, vol. 4, no. 10, Oct. 2025, doi: 10.1016/J.JACADV.2025.102141/SUPPL_FILE/MMC1.PDF.
19. S. Sathyanarayanan, S. Murthy K, C. Gudada, S. Mallappa, and N. Ail, "Heart Sound Analysis with Machine Learning Using Audio Features for Detecting Heart Diseases," International Journal of Computer Information Systems and Industrial Management Applications, vol. 16, no. 2, pp. 131–147, 2024, [Online]. Available: https://cspub-ijcisim.org/index.php/ijcisim/article/view/628
20. S. Sathyanarayanan, S. M. K., C. Gudada, and S. K. Mallappa, "Use Of Audio Transfer Learning To Analyse Heart Sounds For Detecting Heart Diseases," Nanotechnol. Percept., vol. 20, no. S11 (2024), pp. 1863–1878, 2024, doi: http://dx.doi.org/10.62441/nano-ntp.v20iS11.128.
21. S. Sathyanarayanan, S. Murthy Krishnamurthy, C. Gudada, S. Kumar Mallappa, and N. Ail, "Heart Sound Analysis with Machine Learning Using Audio Features for Detecting Heart Diseases," 2024.
22. S. Sathyanarayanan, S. Murthy, S. Mallappa, and C. Gudada, "Machine Learning Approach Using HOG and LBP Features of Spectrograms-based Heart Sounds Analysis for the Detection of Heart Diseases - Voulme 1 - Smart healthcare," in Proceedings of the 15th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2023), A. Bajaj, A. Abraham, and O. Castillo, Eds., London: Springer, Apr. 2025.
23. V. K. Prasad et al., "Revolutionizing healthcare: a comparative insight into deep learning's role in medical imaging," Scientific Reports 2024 14:1, vol. 14, no. 1, pp. 30273-, Dec. 2024, doi: 10.1038/s41598-024-71358-7.
24. S. A. Khalid, T. Khaliq, Y. N. Rehman, Z. Anwar, and W. Mirani, "Comparative Performance of Artificial Intelligence and Radiologists in Detecting Lung Nodules and Breast Lesions on CT and MRI: A Systematic Review," Cureus, vol. 17, no. 11, p. e95943, Nov. 2025, doi: 10.7759/CUREUS.95943.
25. R. El-Khoury and G. Zaatari, "The Rise of AI-Assisted Diagnosis: Will Pathologists Be Partners or Bystanders?," Diagnostics 2025, Vol. 15, Page 2308, vol. 15, no. 18, p. 2308, Sep. 2025, doi: 10.3390/DIAGNOSTICS15182308.
26. "Working groups | International Medical Device Regulators Forum." Accessed: Jul. 10, 2026. [Online]. Available: https://www.imdrf.org/working-groups
27. B. M. Rahamtalla et al., "The AI-Powered Healthcare Ecosystem: Bridging the Chasm Between Technical Validation and Systemic Integration—A Systematic Review," Future Internet 2025, Vol. 17, Page 550, vol. 17, no. 12, p. 550, Nov. 2025, doi: 10.3390/FI17120550.
28. A. A. Armoundas et al., "Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association," Circulation, vol. 149, no. 14, pp. E1028–E1050, Apr. 2024, doi: 10.1161/CIR.0000000000001201/SUPPL_FILE/ARMOUNDAS_AI_SUPPLEMENT.PDF.
29. D. Beals, L. Simon, F. Rogers, and S. Pogroszewski, "Revolutionizing Diabetic Retinopathy Screening: Integrating AI-Based Retinal Imaging in Primary Care," Journal of CME, vol. 14, no. 1, p. 2437294, Dec. 2025, doi: 10.1080/28338073.2024.2437294.
30. W. Bulten et al., "Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists," Modern Pathology, vol. 34, no. 3, p. 660, Mar. 2020, doi: 10.1038/S41379-020-0640-Y.
31. E. Strickland, "IBM Watson, Heal Thyself," Spectrum.IEEE.ORG, 2019.
32. Y. Ma, S. Jamdade, L. Konduri, and H. Sailem, "AI in Histopathology Explorer for comprehensive analysis of the evolving AI landscape in histopathology," npj Digital Medicine 2025 8:1, vol. 8, no. 1, pp. 156-, Mar. 2025, doi: 10.1038/s41746-025-01524-2.
33. L. Hong, M. Teng, M. He, J. Zhao, and S. Li, "AI-driven multi-scale target analysis of traditional Chinese medicine: From the pharmacological effects of single compounds to the synergistic mechanisms of formulae," J. Adv. Res., Mar. 2026, doi: 10.1016/J.JARE.2026.03.050.
34. I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," Nature Medicine 2021 27:10, vol. 27, no. 10, pp. 1735–1743, Sep. 2021, doi: 10.1038/s41591-021-01506-3.

35. M. Li, P. Xu, J. Hu, Z. Tang, and G. Yang, “From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare,” Med. Image Anal., vol. 101, p. 103497, Apr. 2025, doi: 10.1016/J.MEDIA.2025.103497.
36. G. Asti et al., “Development, implementation, and validation of an open-source Federated Learning platform to accelerate innovation and boost personalized medicine in rare and ultra-rare haematological diseases: an initiative by GenoMed4All Consortium,” medRxiv, p. 2025.08.07.25333044, Aug. 2025, doi: 10.1101/2025.08.07.25333044.