JSUCIT

Journal of Shaqra University for Computing and Information Technology
ISSN 2961-4279 (Print) || ISSN 2961-4287 (Online)

From Research Acceleration to Deployment Stagnation: A Critical Review of Post-COVID AI and Machine Learning in Healthcare

Rim Hamdaoui[1*, 2]

[1]Department of Computer Science, College of Science and Humanities, Dawadmi, Shaqra University, Saudi Arabia.

[2]MACS Laboratory: Modeling, Analysis and Control of Systems, National Engineering School of Gabes, Gabes University, Tunisia

r.hamdaoui@su.edu.sa

Abstract The COVID-19 pandemic catalyzed an unprecedented expansion of research in healthcare artificial intelligence (AI) and machine learning (ML), accelerating advances in model architectures, representation learning, and multimodal systems. Despite this rapid methodological progress, the post-pandemic translation of AI/ML research into sustained, large-scale clinical deployment has remained limited. This review critically examines the widening gap between algorithmic innovation and the real-world implementation of healthcare AI/ML from 2019 to 2025. Across several medical domains, we identify recurring failure modes that undermine deployment readiness, including an overreliance on retrospective benchmarks, inadequate external and prospective validation, poor calibration, and vulnerability to dataset shift and distributional drift. We further analyze how emerging paradigms, such as continuously learning and foundation-scale models, challenge existing assumptions around model governance, traceability, and regulatory oversight. By contrasting domain-specific adoption patterns and examining case studies of both successful and stagnant deployments, we show that clinical impact depends less on algorithmic sophistication than on alignment with data stability, workflow integration, and regulatory feasibility. We argue that closing the research–deployment gap requires a shift toward robustness, calibration, and lifecycle-centric AI/ML design, supported by evaluation frameworks that reflect real clinical constraints rather than idealized experimental conditions.

***Index Terms*— Artificial intelligence, machine learning, healthcare, clinical deployment, medical imaging, clinical decision support, model generalization, regulatory challenges.**

I. INTRODUCTION

Since 2019, and particularly following the COVID-19 pandemic, research in artificial intelligence (AI) and machine learning (ML) for healthcare has expanded at an extraordinary rate. Through medical imaging, clinical decision support, triage optimization, predictive modeling, and multimodal data analysis, a rapidly growing body of literature has reported substantial performance gains using advanced AI/ML techniques [1]. This acceleration was fueled by increased availability of digital health data, improvements in computational infrastructure, and urgent clinical demands during the pandemic. These factors have created widespread expectations that AI/ML systems would get a rapid transition from experimental tools to routine components of clinical care [2]. The volume of AI/ML-related publications in medicine has more than tripled in the post-COVID period, spanning a wide range of diseases, clinical tasks, and algorithmic approaches [2], [3]. Fig.1 illustrates the quick expansion of research activity following the COVID-19 pandemic.

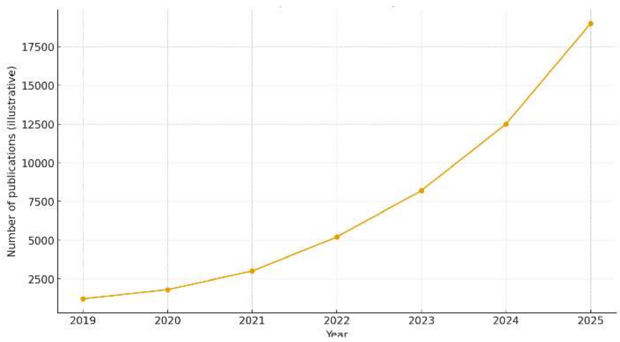

Fig. 1 Growth of AI/ML-related publications in medicine (2019–2025)

Deep learning models have demonstrated impressive results in image-based diagnosis, risk stratification, and outcome prediction, while natural language processing and multimodal learning have extended AI/ML capabilities to unstructured clinical data and longitudinal patient records.

On the surface, these developments suggest a field progressing rapidly toward clinical maturity. But in practice, the translation of these advanced methodologies into real-world clinical implementation has been limited. Despite the successful scientific output, just a small part of published AI/ML systems have progressed to regulatory approval, and few others still have achieved durable integration into real healthcare deployment. Many models that perform well in retrospective, single-center evaluations fail to generalize across institutions, patient populations, or evolving clinical environments [1].

Although methodological innovation has accelerated rapidly, clinical implementation has progressed far more slowly. This disconnection raises a fundamental concern in healthcare AI/ML. In effect, the field has become highly proficient at demonstrating algorithmic potential under controlled conditions, while struggling to ensure reliability, accountability, and safety in real clinical settings [1]–[5]. The causes of this gap are multifactorial. Technically speaking, many healthcare ML models are applied to curated datasets that miss heterogeneity and clinical practice efficiency. This excessive rely on retrospective benchmarks leads to limited potential evaluation. Besides, the rise of continuously learning models and large-scale multimodal architectures challenge regulatory assumptions around version control. These paradigms destabilize the deployment frameworks on which current clinical translation depends [1]. More than technical limitations, non-algorithmic constraints play a decisive role. Indeed, regulatory uncertainty, medico-legal accountability, data governance requirements and limited interoperability with existing electronic health record (EHR) infrastructures continue to slow adoption. Moreover, medical trust and workflow integration remain critical barriers as AI/ML systems function as “black boxes” rather than transparent, assistive tools [2], [3]. These factors reinforce a structural misalignment between academic research incentives and the requirements of real-world clinical implementation.

This review examines and discusses the gap between theoretical advances and practical deployment of AI/ML in healthcare from 2019 to 2025. Rather than merely cataloging algorithmic developments, we analyze recurring failure modes across major application domains; including medical imaging, EHR-based prediction, and multimodal systems and contrast systems that have achieved sustained clinical adoption with those that have stagnated. By doing so, we aim to identify the technical, operational, and regulatory characteristics that distinguish deployable healthcare AI/ML from models that remain confined to the research literature.

As AI/ML researcher, we think it is crucial to investigate the gap between research and real-world deployment because understanding this gap can help identify practical barriers, improve translational strategies, and ultimately ensure that AI/ML innovations lead to tangible improvements in patient outcomes.

This document is structured as follows: The second section describes the methodology of how studies cited in the work, were selected and compared. The third part illustrates the capabilities and deployment of AI/ML in healthcare. The fourth part gives statistics, comparisons and a bibliographic review categorized according to the major medicine fields using intelligent techniques. The gap between the theoretical aspects and the real practical use of AI/ML in healthcare is discussed in the fifth section, including limitations and challenges. Section VI gives a critical evaluation of the Research–Deployment gap and suggests some solutions. Section VII offers a practical value by detailing a workflow-aware AI deployment within EHR environments. The final part concludes the work.

II. Methodology for Analyzing AI/ML Deployment in Healthcare

This review follows a structured approach to examine how healthcare AI/ML systems progressed, or failed to progress, from research prototypes to real-world clinical tools between 2019 and 2025. The aim is to identify recurring translational failure modes and socio-technical patterns that differentiate research acceleration from durable clinical impact.

A targeted literature search was conducted across major scientific databases, including PubMed, Scopus, IEEE Explore, and Web of Science. Search terms combined methodological and translational keywords, including: “artificial intelligence,” “machine learning,” “clinical deployment,” “external validation,” “prospective study,” “regulatory approval,” “FDA,” “workflow integration,” “dataset shift,” and “post-market monitoring.”

To reflect the temporal shift initiated by COVID-19, only studies published between January 2019 and June 2025 were considered. This timeframe captures both the acceleration in AI/ML research output and subsequent attempts at clinical translation.

1. *Inclusion and Exclusion Criteria*

Studies were included if they met at least one of the following criteria:

- Reported real-world deployment, regulatory approval, or operational integration of an AI/ML system in clinical practice.
- Presented prospective validation, multi-center external validation, or post-market performance monitoring.
- Explicitly discussed implementation barriers, governance constraints, or regulatory frameworks affecting AI/ML deployment.
- Provided comparative performance data across institutions or time periods illustrating generalization behavior.

Studies were excluded if they:

- Reported only retrospective, single-center performance without discussion of external validation or translational implications.
- Focused purely on algorithmic novelty without clinical context.
- Were limited to simulation-only experiments with no real-world applicability.
- Lacked sufficient methodological detail to assess deployment relevance.

2. *Screening and Categorization Process*

Selected articles were categorized according to:

- Clinical domain (oncology, cardiology…).
- Model class (CNNs, RNNs, Random Forests, NLP, RL…).
- Validation design (retrospective, external, prospective).
- Regulatory status (research prototype, FDA-cleared, CE-marked, operationally deployed).

3. *Comparative Analytical Framework*

To systematically analyze the research–deployment gap, included studies were evaluated across five dimensions: **validation robustness**, **generalization and stability**, **regulatory alignment**, **workflow integration** and **lifecycle governance.**
Rather than ranking systems solely by accuracy metrics, this framework prioritizes deployment-readiness characteristics. This lens explains why imaging-based screening tools in ophthalmology and cardiology achieved sustained adoption, while many high-performing retrospective models remained confined to the research literature.

III. AI/ML IN MEDICINE: CAPABILITIES AND DEPLOYMENT IMPLICATIONS

A wide range of AI/ML techniques have been applied to healthcare tasks over the last decade, with particularly rapid expansion after 2019. These methods include image-based analysis, sequential modeling of longitudinal data, structured prediction from electronic health records (EHRs), and interpretation of unstructured clinical text.

This section reviews the dominant AI/ML techniques used in healthcare, not from a methodological tutorial perspective, but through the lens of their practical behavior under clinical deployment constraints.

1. Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) is the most known approach for medical image analysis, including most state-of-the-art systems in radiology, pathology and ophthalmology. They have the ability to automatically extract hierarchical features from imaging data giving as a result high-performance detection, classification and segmentation thru modalities such as X-ray, CT, MRI and fundus photography. CNN-based systems have reported near-expert or expert-level accuracy in controlled evaluations, particularly for well-defined imaging tasks. However, the deployment experience has revealed important limitations. Failures observed in real-world settings are often linked to CNN sensitivity. While the innovation in theories and ideas is extremely fast, the implementation is going very slow to dataset composition, acquisition protocols and imaging devices. In fact, models trained on curated, single-institution datasets often show performance degradation when applied to images that are acquired using different scanners, populations or clinical workflows. In addition, CNNs usually provide calibrated uncertainty estimates, limiting their suitability for high-risk diagnostic decisions. CNNs have achieved the most successful clinical deployments among AI/ML techniques, although the challenges [6], [7]. This explains why adoption has concentrated in ophthalmology and radiology rather than in more heterogeneous clinical domains.

2. Recurrent Neural Networks (RNNs)

Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) architectures, are designed to model temporal dependencies in sequential data. In healthcare, these models have demonstrated strong predictive performance for adverse event detection, disease progression, and patient outcome forecasting in chart review studies. They have been widely applied to longitudinal patient records, vital signs, electrocardiograms (ECGs), and time-series data extracted from EHR systems [8]. However, in practice, RNN- and LSTM-based systems face substantial deployment barriers including temporal clinical data irregularity, incomplete, and documentation bias. These factors lead to instability when models are transferred across institutions.

3. Random Forests (RF)

Random Forests (RFs) are ensemble learning methods that combine multiple decision trees to improve predictive robustness and reduce overfitting. They are commonly applied to structured clinical data, including laboratory values, demographic variables, and EHR-derived features [9]. RF models are often favored for their relative interpretability and stable performance on tabular data.

From a deployment perspective, RFs offer advantages over deep neural networks in transparency and simplicity of validation. Feature importance measures provide clinicians with intuitive insights about model behavior, supporting reliance and interpretability. But, RF performance remains constrained by data quality and feature engineering choices. Like other supervised models, RFs trained on institution-specific data may not generalize reliably without extensive external validation, limiting their scalability within healthcare systems.

4. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are widely used in medical classification tasks, mostly in settings with high-dimensional but relatively small datasets, such as genomic and proteomic analysis [10]. Their ability to construct optimal decision boundaries has enabled strong performance in cancer subtype classification and biomarker-based prediction.

However, SVMs face practical limitations in large-scale deployment. In fact, Kernel selection, parameter tuning and limited interpretability complicate integration into clinical workflows. Moreover, SVMs do not naturally accommodate evolving data distributions or continuous learning paradigms, what makes their suitability for dynamic clinical environments less.

5. Natural Language Processing (NLP)

The role of Natural language processing (NLP) techniques is to enable the extraction of clinically relevant information from unstructured text, including physician notes, discharge summaries, and radiology reports [11]. NLP-based models support tasks like named entity recognition, phenotyping, and risk stratification, expanding AI applicability beyond structured data. There are many challenges for NLP systems deployment. Clinical language varies across institutions, specialties, and individual practitioners that lead to significant domain shift. Documentation practices change over time, and delicate linguistic differences can degrade model performance. All these issues, combined with concerns around bias and data governance, have limited the considerable adoption of NLP-driven clinical decision support, despite strong research results.

6. Reinforcement Learning (RL)

Reinforcement learning (RL) appeared as a promising framework for treatment optimization, personalized therapy, sepsis management and dosing strategies, where decisions unfold over time. RL models have demonstrated potential in cited areas, however, they face some of the most significant barriers to real-world deployment. Training requires large volumes of high-quality, temporally consistent data, and learned policies are difficult to validate prospectively. Ethical concerns, safety guarantees, and the challenge of interpreting learned strategies in clinical contexts have confined RL applications to experimental settings.

To summarize, across these methods, a clear pattern appears. Techniques that operate on standardized data modalities, support conservative validation strategies, and integrate smoothly into existing workflows; most notably CNNs in imaging have achieved the greatest deployment success. In contrast, models that rely on heterogeneous, longitudinal, or unstructured data face compounded challenges related to generalization, interpretability, and governance. These observations directly inform the deployment patterns analyzed in Section IV and underscore that algorithmic performance alone is insufficient to predict real-world clinical impact.

IV. DOMAIN-SPECIFIC PATTERNS OF AI/ML ADOPTION IN HEALTHCARE

After COVID'19, AI/ML have been applied across a wide range of medical specialties, yielding several proof-of-concept systems and, in a few cases, sustained clinical deployments. However, adoption has been highly uneven across domains. Differences in data structure, task definition, workflow integration, and regulatory exposure have strongly influenced whether AI/ML systems transitioned from research prototypes to operational tools. This section reviews major clinical domains not to enumerate algorithmic success, but to analyze how domain-specific constraints shape deployment outcomes.

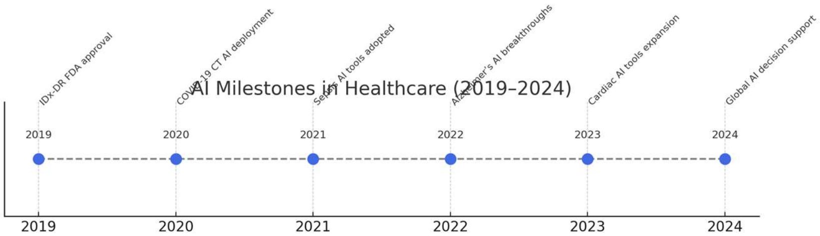

Fig. 2 Timeline of major AI milestones in healthcare since 2019

TABLE 1. STRUCTURED SUMMARY OF REPRESENTATIVE STUDIES REVIEWED (2019–2025)

Study / System	Clinical Domain	Model Type	Validation Design	Deployment Status	Key Deployment Enabler or Barrier
IDx-DR (Autonomous DR Screening)	Ophthalmology	CNN-based imaging	Prospective, multi-site clinical trial	FDA-cleared, real-world deployment	Narrow task scope, standardized input, workflow integration
Viz.ai Stroke Detection	Neurology	Deep learning imaging + workflow routing	Multi-center validation + real-world metrics	Operational in stroke centers	Time-critical use case, seamless PACS integration
Caption Health (AI-guided Ultrasound)	Cardiology	DL-based acquisition guidance	Prospective usability validation	FDA-cleared assistive tool	Human-in-the-loop design, assistive positioning
AI ECG Detection Models	Cardiology	Deep neural networks	External validation across datasets	Partial clinical integration	Standardized signal input, actionable endpoint
Chest X-ray Classifiers (Academic Prototypes)	Pulmonology / Radiology	CNN	Retrospective, single-center	Limited real-world adoption	Poor generalization, lack of prospective validation
Sepsis Prediction Models (EHR-based)	Critical Care	ML / RNN / Gradient Boosting	Retrospective + limited external validation	Limited / decision-support only	Alarm fatigue, distribution shift, unclear accountability
Multimodal Foundation Models	Multi-domain	Transformer-based multimodal architectures	Mostly retrospective benchmarking	Experimental / no routine deployment	Governance complexity, opacity, regulatory uncertainty

The timeline in Fig.2, summarizes the key methodological, regulatory, and deployment-related milestones in healthcare AI since 2019. It highlights the temporal divergence between rapid algorithmic innovation and more gradual progress in validation and clinical integration.

Table.1 illustrates a comparative summary of representative AI/ML systems reviewed, highlighting validation design, deployment status, and structural enablers or barriers influencing real-world integration. Domain-dependent differences in task structure and data stability, partially, explain heterogeneous deployment outcomes.

1. AI/ML in Oncology

Oncology has been one of the most intensively studied areas in healthcare AI, such as cancer detection, classification, and prognosis. ML models, including support vector machines and deep neural networks demonstrated high accuracy in breast cancer diagnosis using curated datasets such as the Wisconsin Breast Cancer dataset, with reported accuracies exceeding 97% [13]. Similar performance has been reported for decision-support systems such as IBM Watson, which aggregate heterogeneous clinical and genomic data [1], [2], [5]. Despite these results, real-world deployment in oncology remains constrained. First of all, because many models are trained on narrowly defined datasets with limited diversity, raising issues about generalizability. Secondly, oncology decision-making often involves multidisciplinary, complex workflows and high medico-legal stakes, which limit tolerance for opaque or weakly validated systems. As a result, oncology AI/ML has largely remained confined to decision support rather than autonomous deployment.

2. AI/ML in Neurology and Alzheimer's disease

In neurology, AI/ML systems are applied to neuroimaging, electrophysiology and longitudinal clinical data, with particular emphasis on early detection of neurodegenerative disorders [14], [15]. Deep learning models trained on MRI and PET imaging have shown encouraging results in distinguishing Alzheimer's disease (AD), Mild Cognitive Impairment (MCI) and healthy aging [16], [17]. Multimodal approaches combining imaging, clinical, and language-based features better improved predictive performance [18], [19]. However, deployment in neurological care remains limited. Neurodegenerative diseases evolve slowly, ground truth labels are often uncertain and longitudinal validation requires long follow-up periods. Besides, imaging protocols and cognitive assessments vary substantially across institutions, amplifying dataset shift.

These factors complicate prospective validation and regulatory approval.

3. AI/ML in Cardiology

Cardiology represents one of the more successful domains for AI/ML deployment; particularly in imaging and signal analysis. FDA-approved systems such as Artery's Cardio DL demonstrated reliable performance in automating ventricular segmentation from cardiac MRI and integrating efficiently into existing workflows [1].

Deep Learning models have also achieved cardiologist-level accuracy in detecting left ventricular dysfunction from electrocardiograms and echocardiograms [20], [21]. Beyond imaging, ML-based risk prediction models using EHR data have shown improved performance over traditional risk scores for cardiovascular events [22], [23]. Wearable-device–based AI systems for arrhythmia detection, exemplified by large-scale studies such as the Apple Heart Study, illustrate successful deployment when tasks are well-defined and outcomes are actionable [24]. Cardiology's relative success reflects standardized data acquisition, clear clinical endpoints, and direct links between predictions and interventions conditions; that decreases deployment barriers.

4. AI/ML in Ophthalmology

Ophthalmology can be considered as the clearest example of sustained AI/ML deployment in healthcare. Indeed, autonomous and semi-autonomous systems for diabetic retinopathy screening, such as IDx-DR and EyeArt, achieved regulatory approval and real-world adoption [25]. Deep Learning models developed by academic and industrial groups demonstrated comparable performance to expert ophthalmologists in detecting diabetic retinopathy, glaucoma, and age-related macular degeneration [26]–[29]. Several factors explain this success; retinal imaging is highly standardized, disease labels are relatively unambiguous and screening workflows are well suited to automation. AI/ML systems in ophthalmology, generally, address high-volume screening tasks rather than complex diagnostic decisions, which reduces medico-legal risk. As a result, ophthalmology serves as a reference case for how narrow task definition and deployment-aware validation enable clinical adoption [30], [31].

5. AI/ML in Pulmonology

AI/ML applications in pulmonology, focused primarily on imaging-based diagnosis of lung diseases, including lung cancer, pneumonia, pulmonary fibrosis, and COVID-19-related abnormalities [32], [33]. Deep Learning models achieved high accuracy in controlled evaluations as AI/ML systems outperformed clinicians in specific tasks like pulmonary function test interpretation [34]. Nevertheless, deployment remains inconsistent due to sensitivity to equipment of pulmonary imaging, acquisition protocols and patient cooperation. These reasons introduce variability that challenges model robustness. Predictive models for disease progression in conditions like COPD and idiopathic pulmonary fibrosis are promising, but they still need long-term validation in real patients [35]. In intensive care, ML models can help predict mortality and the need for ventilation but they are used only as decision-support tools, not as independent systems [36].

6. AI/ML in Medical Imaging

Across specialties, medical imaging represents the most mature application area for AI/ML deployment. Meta-analyses have shown that Deep Learning models achieve diagnostic performance comparable to healthcare professionals across multiple imaging tasks [37], [38]. These successes, however, should be interpreted in light of strict task boundaries and relatively stable data modalities. Imaging-based AI/ML benefits from standardized inputs, visual interpretability, and well-established validation pathways. In contrast, domains that rely heavily on longitudinal, multimodal or unstructured data face many challenges in relation with generalization and governance. This contrast underscores why imaging dominates deployed healthcare AI/ML while other applications lag behind.

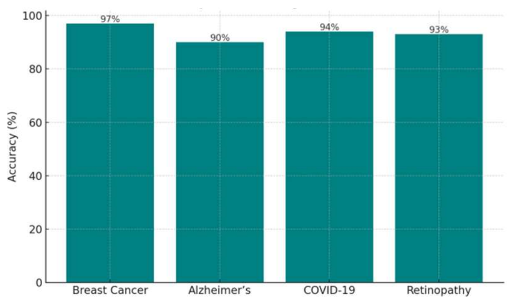

Fig. 3 Comparative diagnostic performance of AI/ML systems across clinical domains

Fig.3 reports diagnostic performance of AI/ML systems across selected disease areas. While high accuracy is frequently achieved in controlled evaluations, these metrics do not necessarily translate into robustness or generalizability under real-world clinical conditions.

These domain-specific patterns, taken together, reinforce an important conclusion: AI/ML systems succeed in healthcare not because of novel algorithms, but because the task is well-defined, the data are stable, and the system fits into existing workflows and regulatory constraints. Fields like ophthalmology and cardiology show strong adoption because they rely on high-volume, standardized tasks with measurable outputs. In contrast, areas with heterogeneous data, complex decision chains, and high uncertainty remain difficult to deploy, even when research models achieve high benchmark performance.

These observations directly inform the deployment gap analyzed in Section V, illustrating that real-world impact depends on socio-technical alignment rather than algorithmic capability alone.

V. DISCUSSION

1. Theoretical Advances versus Practical Achievements

During the past five years, AI/ML have delivered demonstrable gains in healthcare research, especially in diagnostic accuracy, pattern recognition and risk stratification. Across multiple application domains, ML models consistently overtake traditional statistical methods on controlled classification and prediction benchmarks, reporting high accuracy in tasks such as image-based diagnosis and risk stratification [38]–[43]. These systems demonstrate strong pattern recognition capability, fast inference and scalable automation. This indicates that core model performance is no longer the main technical limitation. Instead, the remaining challenges lie in robustness, generalization, and real-world integration.

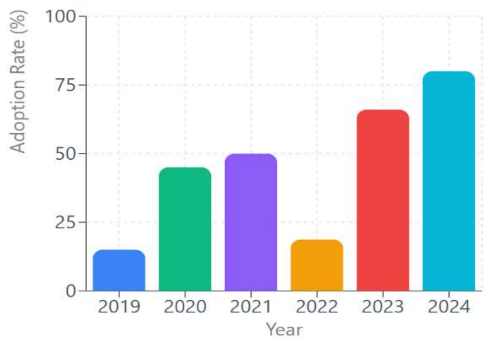

Fig. 4 Adoption of AI/ML technologies in hospital settings (2019–2024)

Fig.4 illustrates increasing institutional interest in AI/ML in hospital-level adoption from 2019 to 2024, including pilot implementations and partial deployments, alongside persistent barriers to full-scale, sustained integration into clinical workflows. These achievements, however, must be interpreted with caution. While headline performance metrics suggest readiness for deployment, they generally hide a critical reality; most results are obtained under narrowly defined experimental conditions that do not reflect the complexity of real clinical environments. Reported gains are frequently derived from retrospective datasets, curated cohorts, and single-institution studies, where data distributions are stable, labels are carefully constructed, and operational constraints are largely absent. When exposed to heterogeneous patient populations, evolving clinical practices, or shifts in data acquisition protocols, model performance frequently degrades in clinically meaningful ways [31], [38]. This divergence between theoretical capability and practical reliability has become increasingly visible in post-pandemic. Despite widespread enthusiasm and rising adoption statistics, deployment remains concentrated in a limited set of applications, predominantly imaging-based tasks in ophthalmology and radiology [31]. Predictive models based on EHR data, although abundant in the literature, often underperform when transferred across institutions or time periods, limiting their clinical usefulness [1]. Important challenges such as regulatory, interoperability and medico-legal constraints slow the integration even when technical feasibility appears promising [44]. Besides, black-box models, particularly those lacking transparent uncertainty estimates or interpretable reasoning, face resistance in high-stakes decision-making contexts [1], [45]. In such settings, marginal gains in accuracy are insufficient to justify the loss of responsibility or explainability required for safe clinical practice. All these factors explain why many theoretically advanced systems fail to achieve sustained real-world impact.

2. Structural Limitations and Deployment Barriers

A recurring limitation in healthcare AI/ML research lies in the gap between model development and deployment readiness. Systematic reviews since 2020 consistently report that the majority of published models lack external validation and are rarely evaluated prospectively [46]–[49]. Performance metrics reported in the literature often decline significantly when models are tested on independent healthcare systems, revealing a translation bottleneck between research success and operational viability.

Regulatory processes represent a second major constraint. While agencies such as the FDA, MHRA, and Health Canada have begun to articulate Good Machine Learning Practice (GMLP) principles and predetermined change control plans for adaptive algorithms, implementation remains challenging in practice [50], [51]. Continuous-learning models raise unresolved questions regarding safety monitoring, version control, and accountability. As a result, most AI/ML systems currently approved for clinical use remain “locked” models, limiting their ability to adapt to evolving clinical environments.

Equally, another significant challenge is the operational integration. Many research prototypes are developed in isolation from production constraints and fail to address EHR interoperability, clinical workflow compatibility, user-interface requirements, training pipelines, and post-deployment monitoring infrastructure [52], [53]. As a result, models that perform well in offline benchmarks often fail during deployment. The lack of deployment-aware system design from the earliest development stages remains a primary reason why technically strong algorithms do not translate into real-world systems.

Besides, algorithmic bias represents a major deployment risk, particularly in heterogeneous populations. Differences in demographic representation, socioeconomic status, and healthcare access can produce systematically unequal model performance. Deployment-ready systems should incorporate fairness audits, subgroup performance reporting, and bias monitoring mechanisms during post-market surveillance to ensure equitable clinical impact.

3. Case Studies: Successful versus Stagnant Deployment

A comparison between systems that achieved sustained deployment and those that did not reveals consistent patterns will be presented by here.

A. Successful Deployments

Autonomous diabetic retinopathy screening systems such as IDx-DR succeeded by addressing a narrowly defined clinical task with clear operational boundaries [54], [55]. Their development was supported by prospective, preregistered trials, a well-defined intended use, and seamless integration into primary care workflows. These systems demonstrated immediate clinical and economic value by reducing unnecessary referrals, facilitating regulatory approval and adoption. Similarly, Viz.ai’s stroke detection platform addressed a time-critical workflow blockage, large vessel occlusion triage, where speed directly impacts patient outcomes [56, [57]. Instead of trying to replace clinicians, these systems were designed to fit directly into existing radiology workflows. By proving that they could reduce time-to-treatment in real settings, they showed value not only in model accuracy but also in operational efficiency. Similarly, AI-guided ultrasound tools such as Caption Health focused on helping users acquire better scans rather than making diagnoses [58]. By guiding non-experts in real time and positioning the system as an assistive tool, these platforms will be easier to regulate, easier to use, and more readily accepted by clinicians. This kind of workflow-aware design is a key reason why some AI/ML systems succeed in practice while many high-performing algorithms fail outside the lab.

B. Stagnant or Limited Deployments.

In contrast, many academic imaging models; particularly CNN-based chest X-ray classifiers have struggled to move beyond proof-of-concept [59]. These systems are generally trained on curated datasets, exhibit poor generalization across institutions and lack prospective validation. Without clear reimbursement pathways or workflow integration strategies, high benchmark accuracy alone has proven insufficient for adoption.

EHR-based prediction systems, such as sepsis risk models, face even greater deployment challenges [60]. Their performance often drops when used in new hospitals or patient populations. Frequent false alerts overwhelm clinicians and lead to alarm fatigue. At the same time, many models provide little explanation for their predictions, and responsibility for errors remains unclear. Uncertain regulations around continuously updating risk models make deployment even harder. Multimodal foundation models represent an emerging but particularly complex case. While technically powerful, their immense data and computational requirements, limited explainability, and unresolved ethical and regulatory questions have prevented real-time clinical deployment [61].

VI. Critical Evaluation and Practical Pathways

1. Critical Evaluation of the Research–Deployment gap

Despite the maturation of healthcare AI/ML research since 2019, the gap between published performance claims and systems that measurably improve patient outcomes in routine care remains substantial. This gap is not primarily the result of insufficient algorithmic capability. Instead, it reflects structural, organizational, and human-centered limitations that are systematically underestimated in research settings. A central issue is the persistent reliance on idealized data environments. Many healthcare ML models are developed using curated, retrospectively labeled datasets that fail to reflect the heterogeneity, incompleteness, and temporal instability of real clinical data. Systematic reviews consistently demonstrate that most published models lack robust external validation and that reported performance often degrades significantly when evaluated across institutions or time periods [46]–[49]. This disconnect reveals a fundamental mismatch between research conditions and deployment realities.

Regulatory and validation barriers further intensify this divide. Although regulatory bodies have begun to articulate Good Machine Learning Practice (GMLP) principles and frameworks for adaptive algorithms, translating these guidelines into operational pathways remains challenging [50], [51]. Continuous-learning systems, in particular, raise unresolved questions related to safety monitoring, version control, and accountability. As a result, most clinically deployed AI/ML systems remain static, "locked" models, with limited ability to adapt to clinical environments.

Equally critical is the gap between model development and operational integration. Many research prototypes are not designed with deployment constraints in mind, overlooking essential factors such as interoperability with EHR systems, clinician workflows, user-interface design, staff training and post-deployment monitoring [52], [53]. In practice, the transition from a high-performing research model to a clinically viable system requires substantial engineering, organizational alignment and long-term maintenance efforts that are rarely taken into consideration in academic research.

To conclude, we can say that culture and trust matter just as much as technical performance. Clinicians are understandably cautious about using black-box models whose decisions cannot be clearly explained or questioned. In high-risk settings, small gains in prediction accuracy are not worth losing transparency, responsibility, or professional control. As a result, opaque AI/ML systems are still difficult to adopt, especially when they directly affect treatment decisions rather than supporting screening or workflow efficiency. Taken together, this shows that real-world deployment is not keeping pace with the rapid growth of AI/ML research and hype. Rather, the post-COVID period has primarily clarified the nature of the gap, revealing healthcare AI/ML as a socio-technical challenge that cannot be resolved through algorithmic innovation alone.

Even thought, to mitigate this concern related to opaque "black-box" systems, several explainable AI (XAI) techniques have been adopted in clinical contexts. Post-hoc interpretability methods such as SHAP (Shapley Additive Explanations), LIME, saliency maps in imaging models, and attention-weight visualization in transformer architectures provide local and global interpretability [62].

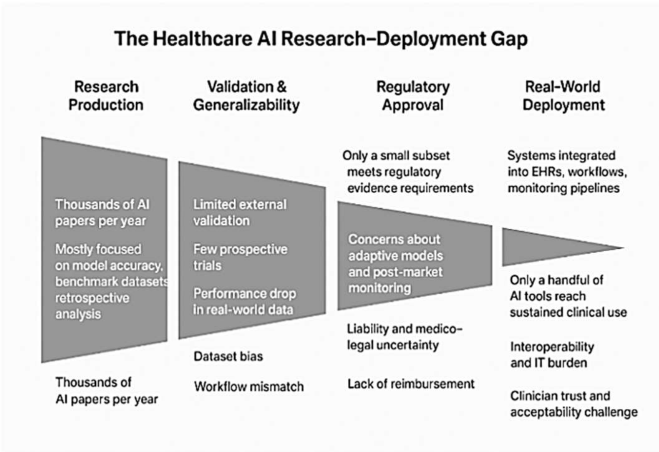

Fig. 5 The healthcare AI/ML research–deployment gap

Fig.5 emphasizes that high research volume does not directly translate into high deployment impact due to technical, regulatory, and operational constraints.

2. Future Practical Pathways of AI/ML Deployment

Closing the gap between theory and practice requires a change in priorities across the healthcare AI/ML ecosystem, encompassing researchers, developers, healthcare institutions and regulators.

From a research perspective, deployment considerations must be integrated early rather than treated as downstream concerns. External validation across heterogeneous populations, prospective evaluation and robustness to dataset shift should be viewed as baseline requirements rather than optional extensions. Models should be designed with calibration, uncertainty estimation, and failure detection in mind, reflecting the realities of clinical decision-making under uncertainty.

For developers and vendors, investment in lifecycle management is essential. This includes systematic monitoring of model performance after deployment, rigorous documentation of data provenance and model updates and transparent change-control mechanisms aligned with regulatory expectations. Multi-site studies conducted prior to commercialization can reduce downstream deployment risk and strengthen the evidence base required for adoption. Besides, reproducibility must be treated as a foundational requirement for deployment readiness. Beyond reporting retrospective performance metrics, developers should ensure transparent documentation of data provenance, model architecture, preprocessing pipelines, and evaluation protocols.

Hospitals, insurers, and healthcare organizations also have a major influence on whether AI/ML systems succeed in practice. They should not judge systems only by retrospective accuracy scores, but by whether they actually improve clinical outcomes, fit into daily workflows, and reduce costs. Clear rules for model monitoring, maintenance, and responsibility are essential for safe and long-term use. Without this institutional readiness, even well-designed AI tools are unlikely to have real impact.

Regulators and policymakers also need to turn high-level principles into practical rules. This means defining clear evaluation standards, audit procedures, and post-deployment monitoring, especially as adaptive and large-scale models become more common. Clear regulation does not slow innovation, it makes responsible deployment possible by reducing uncertainty for developers and healthcare providers.

Robust version control mechanisms should include model version identifiers linked to dataset snapshots, preprocessing configurations, and training metadata. Automated safety triggers; such as performance degradation thresholds, calibration drift alerts, or out-of-distribution detection flags, should be integrated into monitoring pipelines to prevent silent model failure. These mechanisms align with emerging Good Machine Learning Practice (GMLP) principles and facilitate regulatory compliance.

To summarize, progress depends on aligning incentives across all stakeholders. The current focus on benchmark results and publication novelty must be balanced with the real effort required to deploy, validate, and maintain AI/ML systems in clinical environments. The most effective path forward is early collaboration between clinicians, data scientists, engineers, regulators, and ethicists, working together from the start.

VII. WORKFLOW-AWARE AI DESIGN: EHR ENVIRONMENTS

A recurring theme throughout this review is that the failure of many healthcare AI systems does not come from insufficient predictive performance, but from inadequate integration into clinical practice. The electronic health record (EHR) is the central operational backbone of modern healthcare, and any AI system. Consequently, workflow-aware AI design represents a critical bridge between algorithmic capability and real-world deployment.

This section examines how embedding AI systems within EHR environments, through important layers that transforms predictive models from experimental tools into operational clinical infrastructure.

Effective integration operates across four coordinated layers: **automated data extraction**, **event-driven inference**, **native interface embedding**, and **continuous monitoring** as presented in fig. 6.

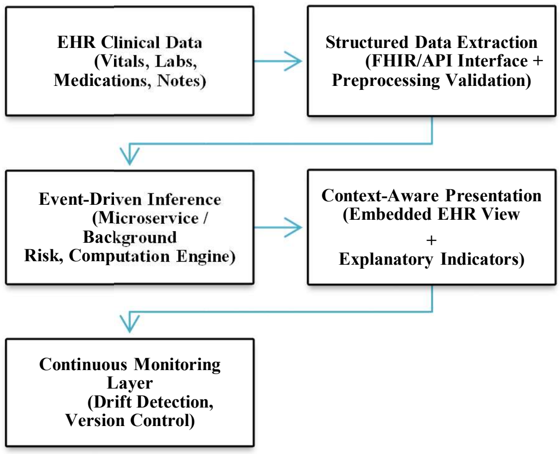

Fig. 6 Conceptual architecture of workflow-aware AI deployment within EHR environments.

Layer 1- EHR Clinical Data: The pathway begins with routinely collected clinical data inside the EHR: vital signs, laboratory values, medications, diagnoses, and timestamps. These are not research-curated datasets but live, heterogeneous inputs generated during care delivery. The quality and structure of this data directly determine downstream model reliability. Through standardized interfaces such as Fast Healthcare Interoperability Resources FHIR-based APIs, relevant variables are programmatically extracted and transformed using preprocessing steps consistent with model training (normalization, missing-value handling, temporal alignment) [63]. This layer ensures that what the model receives in deployment matches what it learned during development.

Layer 2 - Event-Driven Inference: Once structured inputs are available, inference is triggered automatically by clinical events; such as a new lab result or vital sign entry. A background micro service recalculates risk in real time without manual activation. This preserves temporal relevance and embeds AI within the clinical data flow rather than as a parallel tool.

Layer 3 - Context-Aware Presentation: The computed output is not shown as an isolated probability. It is embedded within the clinician's existing EHR interface, accompanied by calibrated risk levels, trend information, and concise explanatory indicators. Output design directly influences whether predictive systems are integrated into clinical decision-making or dismissed as operational noise.

Layer 4 - Continuous Monitoring and Governance: After deployment, performance must be supervised. Calibration drift, subgroup performance, alert frequency, and distributional changes are continuously tracked. Review mode will be activated by automated safety triggers if predefined degradation thresholds are exceeded. Version control ensures traceability of updates. Most research papers stop at the step of model training and validation. However, deployment demands ongoing supervision, performance accountability and regulatory traceability. Layer 4 turns a static algorithm into a governed clinical system.

VIII. CONCLUSION

Over the past six years, healthcare AI/ML have achieved remarkable progress at the methodological level, particularly in imaging-based diagnosis, predictive modeling and multimodal data integration. However, this review demonstrates that technical advancement alone has not translated into widespread, durable clinical deployment. Instead, the post-COVID period has exposed a persistent and structural gap between research performance and real-world impact. The evidence reviewed here indicates that the deployment gap is not primarily an algorithmic problem. In fact, it reflects misalignment between research incentives, regulatory processes, healthcare infrastructure and clinical practice.

Finally, the future impact of AI/ML in healthcare will not be determined by increasingly complex models alone, but by the field's ability to align technical innovation with the realities of clinical care. Without this alignment, continued research acceleration risks deepening the divide between theoretical promise and practical benefit.

REFERENCES

[1] J. A. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: A practical introduction," BMC Medical Research Methodology, vol. 19, no. 1, Mar. 2019. doi:10.1186/s12874-019-0681-4

[2] L. Li et al., "Enhancing lung cancer detection through hybrid features and machine learning hyperparameters optimization techniques," Heliyon, vol. 10, no. 4, Feb. 2024. doi:10.1016/j.heliyon.2024.e26192

[3] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based Deep Learning," Cell, vol. 172, no. 5, Feb. 2018. doi:10.1016/j.cell.2018.02.010

[4] R. S. Moorthy and P. Pabitha, "Intelligent Health Care System using modified feature selection algorithm," Lecture Notes in Electrical Engineering, pp. 777–787, 2022, doi:10.1007/978-981-19-1520-8_63

[5] A. İ. Tekkeşin, "Artificial Intelligence in healthcare: Past, present and future," The Anatolian Journal of Cardiology, 2019. doi:10.14744/anatoljcardiol.2019.28661

[6] G. Jha, A. Jha, and E. John, "A high accuracy CNN for breast cancer detection using mammography images," 2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1196–1200, Aug. 2024. doi:10.1109/mwscas60917.2024.10658807

[7] L. Dong and K. Inoue, "Diagnosis of breast cancer from mammogram images based on CNN," Journal of the Institute of Industrial Applications Engineers, vol. 8, no. 4, pp. 117–121, Oct. 2020. doi:10.12792/jiiae.8.117

[8] T. Poongodi, D. Sumathi, P. Suresh, and B. Balusamy, "Deep learning techniques for electronic health record (EHR) analysis," Studies in Computational Intelligence, pp. 73–103, Jul. 2020. doi:10.1007/978-981-15-5495-7_5

[9] K. Sumwiza, C. Twizere, G. Rushingabigwi, P. Bakunzibake, and P. Bamurigire, "Enhanced cardiovascular disease prediction model using random forest algorithm," Informatics in Medicine Unlocked, vol. 41, p. 101316, 2023. doi:10.1016/j.imu.2023.101316

[10] A. Yaqoob and N. K. Verma, "Feature selection in breast cancer gene expression data using KAO and AOA with SVM classification," Journal of Medical Systems, vol. 49, no. 1, Mar. 2025. doi:10.1007/s10916-025-02171-6

[11] S. Wu et al., "Deep learning in clinical natural language processing: A methodical review," Journal of the American Medical Informatics Association, vol. 27, no. 3, pp. 457–470, Dec. 2019. doi:10.1093/jamia/ocz200

[12] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," Nature Medicine, vol. 24, no. 11, pp. 1716–1720, Oct. 2018. doi:10.1038/s41591-018-0213-5

[13] N. Saxena, R. Vaishnav, S. Saxena, and U. Kumar, "Risk prediction and diagnosis of breast cancer using ML algorithms," Combating Women's Health Issues with Machine Learning, pp. 120–135, Aug. 2023. doi:10.1201/9781003378556-8

[14] A. Esteva et al., "A guide to deep learning in Healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, Jan. 2019. doi:10.1038/s41591-018-0316-z

[15] E. J. Topol, "High-performance medicine: The convergence of human and Artificial Intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, Jan. 2019. doi:10.1038/s41591-018-0300-7

[16] R. Zhang, J. Sheng, Q. Zhang, J. Wang, and B. Wang, "A review of Multimodal Fusion–based Deep Learning for Alzheimer's disease," Neuroscience, vol. 576, pp. 80–95, Jun. 2025. doi:10.1016/j.neuroscience.2025.04.035

[17] T. Zhang et Al. "Predicting MCI to AD Conversation Using Integrated sMRI and rs-fMRI: Machine Learning and Graph Theory Approach", Frontiers in Aging Neuroscience, Volume 13 – 2021, https://doi.org/10.3389/fnagi.2021.688926

[18] K. C. Fraser, K. L. Fors, M. Eckerström, F. Öhman, and D. Kokkinakis, "Predicting MCI status from multimodal language data using cascaded classifiers," Frontiers in Aging Neuroscience, vol. 11, p. 205, Aug. 2019, doi: 10.3389/fnagi.2019.00205.

[19] H. Balabin et al., "Natural language processing-based classification of early Alzheimer's disease from connected speech," Alzheimer S & Dementia, vol. 21, no. 2, p. e14530, Jan. 2025, doi: 10.1002/alz.14530.

[20] S. Katsushika et al., "The Effectiveness of a Deep Learning Model to Detect Left Ventricular Systolic Dysfunction from Electrocardiograms," International Heart Journal, vol. 62, no. 6, pp. 1332–1341, Nov. 2021, doi: 10.1536/ihj.21-407.

[21] D. Eng et al., "Automated coronary calcium scoring using deep learning with multicenter external validation," Npj Digital Medicine, vol. 4, no. 1, p. 88, Jun. 2021, doi: 10.1038/s41746-021-00460-1.

[22] C. Krittanawong et al., "Deep learning for cardiovascular medicine: A practical primer," European Heart Journal, Art. no. ehz056, Feb. 2019, doi: 10.1093/eurheartj/ehz056.

[23] S. Bozyel et al., "Artificial Intelligence-Based Clinical Decision Support Systems in cardiovascular diseases," The Anatolian Journal of Cardiology, pp. 74–86, Jan. 2024, doi: 10.14744/anatoljcardiol.2023.3685.

[24] M. V. Perez et al., "Large-Scale assessment of a smartwatch to identify atrial fibrillation," New England Journal of Medicine, vol.

381, no. 20, pp. 1909–1917, Nov. 2019, doi: 10.1056/nejmoa1901183.
[25] Z. Wang, P. A. Keane, M. Chiang, C. Y. Cheung, T. Y. Wong, and D. S. W. Ting, "Artificial intelligence and deep learning in ophthalmology," in Artificial Intelligence in Medicine, 2021, pp. 1–34. doi: 10.1007/978-3-030-58080-3_200-1.
[26] R. Abbasi et al., "Diabetic retinopathy detection using adaptive deep convolutional neural networks on fundus images," Scientific Reports, vol. 15, no. 1, p. 24647, Jul. 2025, doi: 10.1038/s41598-025-09394-0.
[27] H. Kayani, "Artificial intelligence and its applications in ophthalmology," Journal of Fatima Jinnah Medical University, vol. 13, no. 4, Jan. 2020, doi: 10.37018/jfjmu.724.
[28] M. Trinh, M. Ghassibi, and R. Lieberman, "Artificial intelligence in retina," Advances in Ophthalmology and Optometry, vol. 6, pp. 175–185, Jun. 2021, doi: 10.1016/j.yaoo.2021.04.012.
[29] M. J. A. Girard and L. Schmetterer, "Artificial intelligence and deep learning in glaucoma: Current state and future prospects," Progress in Brain Research, vol. 257, pp. 37–64, Jan. 2020, doi: 10.1016/bs.pbr.2020.07.002.
[30] Y. Tong, W. Lu, Q.-Q. Deng, C. Chen, and Y. Shen, "Automated identification of retinopathy of prematurity by image-based deep learning," Eye And Vision, vol. 7, no. 1, p. 40, Aug. 2020, doi: 10.1186/s40662-020-00206-2.
[31] D. S. W. Ting, H. Lin, P. Ruamviboonsuk, T. Y. Wong, and D. A. Sim, "Artificial intelligence, the internet of things, and virtual clinics: ophthalmology at the digital translation forefront," The Lancet Digital Health, vol. 2, no. 1, pp. e8–e9, Dec. 2019, doi: 10.1016/s2589-7500(19)30217-1.
[32] M. J. P. Priyadarsini et al., "Lung diseases detection using various deep learning algorithms," Journal of Healthcare Engineering, vol. 2023, no. 1, p. 3563696, Jan. 2023, doi: 10.1155/2023/3563696.
[33] S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches," Chaos Solitons & Fractals, vol. 140, p. 110170, Jul. 2020, doi: 10.1016/j.chaos.2020.110170.
[34] M. Topalovic et al., "Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests," European Respiratory Journal, vol. 53, no. 4, p. 1801660, Feb. 2019, doi: 10.1183/13993003.01660-2018.
[35] S. L. F. Walsh et al., "Deep learning–based outcome prediction in progressive fibrotic lung disease using High-Resolution Computed Tomography," American Journal of Respiratory and Critical Care Medicine, vol. 206, no. 7, pp. 883–891, Jun. 2022, doi: 10.1164/rccm.202112-2684oc.
[36] A. Vaid et al., "Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation," Journal of Medical Internet Research, vol. 22, no. 11, p. e24018, Oct. 2020, doi: 10.2196/24018.
[37] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," Neurocomputing, vol. 444, pp. 92–110, Jan. 2021, doi: 10.1016/j.neucom.2020.04.157.
[38] X. Liu et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," Lancet Digital Health, vol. Volume 1, no. Issue 6, Art. No 271–297, Oct. 2019.
[39] R. B. A. Baten, "How are US hospitals adopting artificial intelligence? Early evidence from 2022," Health Affairs Scholar, vol. 2, no. 10, p. qxae123, Sep. 2024, doi: 10.1093/haschl/qxae123
[40] N. Sahni, G. Stein, R. Zemmel, and D. Cutler, "The potential impact of artificial intelligence on healthcare spending," Jan. 2023. doi: 10.3386/w30857.
[41] E. G. Poon, C. H. Lemak, J. C. Rojas, J. Guptill, and D. Classen, "Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges," Journal of the American Medical Informatics Association, vol. 32, no. 7, pp. 1093–1100, May 2025, doi: 10.1093/jamia/ocaf065.
[42] N. Ogbonna, "How AI achieves 94% accuracy in early disease detection: new research findings GlobalRPH," GlobalRPH, April. 01, 2025. https://globalrph.com/2025/04/how-ai-achieves-94-accuracy-in-early-disease-detection-new-research-findings/
[43] S. A. Alowais et al., "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," BMC Medical Education, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.
[44] M. Nasir, K. Siddiqui, and S. Ahmed, "Ethical-legal implications of AI-powered healthcare in critical perspective," Frontiers in Artificial Intelligence, vol. 8, p. 1619463, Jul. 2025, doi: 10.3389/frai.2025.1619463.
[45] S. Sablone, M. Bellino, A. N. Cardinale, M. Esposito, F. Sessa, and M. Salerno, "Artificial intelligence in healthcare: an Italian perspective on ethical and medico-legal implications," Frontiers in Medicine, vol. 11, p. 1343456, Jun. 2024, doi: 10.3389/fmed.2024.1343456.
[46] H.-T. Hu et al., "Addressing data heterogeneity in distributed medical imaging with heterosync learning," Nature Communications, vol. 16, no. 1, p. 9416, Oct. 2025, doi: 10.1038/s41467-025-64459-y.
[47] [47] A. C. Yu, B. Mohajer, and J. Eng, "External validation of deep learning Algorithms for radiologic diagnosis: A Systematic review," Radiology Artificial Intelligence, vol. 4, no. 3, p. e210064, May 2022, doi: 10.1148/ryai.210064.
[48] B. S. Kelly et al., "Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE)," European Radiology, vol. 32, no. 11, pp. 7998–8007, Apr. 2022, doi: 10.1007/s00330-022-08784-6.
[49] M. I. Ahmed, B. Spooner, J. Isherwood, M. Lane, E. Orrock, and A. Dennison, "A Systematic review of the barriers to the implementation of artificial intelligence in healthcare," Cureus, vol. 15, no. 10, p. e46454, Oct. 2023, doi: 10.7759/cureus.46454.
[50] T. H. Fakhouri, Q. Liu, and M. K. ElZarrad, "Artificial intelligence in regulatory decision-making for drug and biological products," in Elsevier eBooks, 2023, pp. 373–375. doi: 10.1016/b978-0-443-15688-5.00011-5.
[51] A. Z. A. Meslamani, "Technical and regulatory challenges of digital health implementation in developing countries," Journal of Medical Economics, vol. 26, no. 1, pp. 1057–1060, Aug. 2023, doi: 10.1080/13696998.2023.2249757.
[52] W. E. M. Berkhout et al., "Operationalization of artificial intelligence applications in the intensive care unit," JAMA Network Open, vol. 8, no. 7, p. e2522866, Jul. 2025, doi: 10.1001/jamanetworkopen.2025.22866.
[53] Y. Shin, M. Lee, Y. Lee, K. Kim, and T. Kim, "Artificial Intelligence-Powered Quality Assurance: Transforming diagnostics, surgery, and Patient Care—Innovations, limitations, and future directions," Life, vol. 15, no. 4, p. 654, Apr. 2025, doi: 10.3390/life15040654.
[54] A. Md, L. Pt, B. M, S. N, and F. Jc, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," Yearbook of Pediatric Endocrinology, September. 2019, doi: 10.1530/ey.16.12.1.
[55] P. Bidwai, S. Gite, K. Pahuja, and K. Kotecha, "A Systematic Literature Review on Diabetic Retinopathy using an Artificial Intelligence approach," Big Data and Cognitive Computing, vol. 6, no. 4, p. 152, Dec. 2022, doi: 10.3390/bdcc6040152.
[56] A. E. Sarhan and M. H. Moawad, "Abstract 278: Automated Emergent large vessel Occlusion Detection using Viz.ai Software And its Impact on stroke workflow metrics and patient outcomes in stroke centers: a Systematic review and Meta-analysis," Stroke Vascular and Interventional Neurology, vol. 5, no. S1, Nov. 2025, doi: 10.1161/svi270000_278.
[57] K. Van Orden et al., "(VISIION-S): Viz.ai Implementation of Stroke augmented Intelligence and communications platform to improve

Indicators and Outcomes for a comprehensive stroke center and Network – Sustainability," Journal of Stroke and Cerebrovascular Diseases, vol. 32, no. 10, p. 107303, Aug. 2023, doi: 10.1016/j.jstrokecerebrovasdis.2023.107303.

[58] A. Narang et al., "Utility of a Deep-Learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use," JAMA Cardiology, vol. 6, no. 6, p. 624, Feb. 2021, doi: 10.1001/jamacardio.2021.0185.

[59] R. Channa, R. Wolf, and M. D. Abramoff, "Autonomous artificial intelligence in Diabetic retinopathy: From algorithm to Clinical application," Journal of Diabetes Science and Technology, vol. 15, no. 3, pp. 695–698, Mar. 2020, doi: 10.1177/1932296820909900.

[60] A. H. Van Der Vegt, I. A. Scott, K. Dermawan, R. J. Schnetler, V. R. Kalke, and P. J. Lane, "Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework," Journal of the American Medical Informatics Association, vol. 30, no. 7, pp. 1349–1361, May 2023, doi: 10.1093/jamia/ocad075.

[61] G. Azarfar et al., "Responsible adoption of multimodal artificial intelligence in health care: promises and challenges," The Lancet Digital Health, p. 100917, Dec. 2025, doi: 10.1016/j.landig.2025.1009

[62] M. Saarela and V. Podgorelec, "Recent Applications of Explainable AI (XAI): A Systematic Literature review," Applied Sciences, vol. 14, no. 19, p. 8884, Oct. 2024, doi: 10.3390/app14198884.

[63] C. R. Nwokocha, A. C. Peter-Anyebe, and O. M. Ijiga, "Evaluating FHIR-Driven interoperability frameworks for secure system migration and data exchange in U.S. health information networks," International Journal of Scientific Research in Science and Technology, p. 781, Jul. 2021, doi: 10.32628/ijsrst523105135