Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

ARTIFICIAL INTELLIGENCE AND HEALTH ANALYTICS IN MODERN HEALTHCARE SYSTEMS: A SYSTEMATIC REVIEW OF DIGITAL TRANSFORMATION, DECISION INTELLIGENCE, AND OPERATIONAL RESILIENCE

Afaq Ahmad Khan

BPP University (MSc)

afaqahmad.sr@gmail.com

DOI:https://doi.org/10.5281/zenodo.21851021

Keywords

artificial intelligence; health analytics; digital transformation; decision intelligence; operational resilience; machine learning; electronic health records; explainable AI; large language models; healthcare operations

Article History

Received: 30 January 2026
Accepted: 15 March 2026
Published: 31 March 2026

Copyright @Afaq Ahmad Khan

Corresponding Author:
***Afaq Ahmad Khan.**
BPP University (MSc)
afaqahmad.sr@gmail.com

Abstract

Artificial intelligence (AI) and health analytics are moving from isolated diagnostic tools toward infrastructure for learning health systems. This systematic review synthesizes recent evidence on how AI supports digital transformation, decision intelligence, and operational resilience across modern healthcare. A structured search of PubMed/MEDLINE and supplementary publisher- and index-based searches was conducted, emphasizing peer-reviewed studies published from 2020 through 2026 while retaining a small number of foundational studies from 2018-2019. Search concepts combined artificial intelligence, machine learning, deep learning, natural language processing, large language models, electronic health records, medical imaging, predictive analytics, clinical decision support, digital transformation, patient flow, and operational resilience. Sixty-eight candidate records were assembled, six duplicates were removed, 62 records were screened, 45 reports underwent full-text or complete-abstract eligibility assessment, and 40 core studies were included in the evidence matrix. The synthesis identified six interacting domains: predictive clinical analytics; electronic health record and language intelligence; imaging and multimodal diagnostics; precision, population, and privacy-preserving analytics; operational intelligence; and governance-enabled decision intelligence. Across these domains, AI improved risk stratification, diagnostic interpretation, documentation, triage, admission and discharge forecasting, and resource planning, but gains were frequently constrained by retrospective designs, local data dependence, dataset shift, limited external validation, workflow friction, and uncertainty about accountability. Recent work increasingly reframes high-performing models as only one component of trustworthy deployment, emphasizing fairness, traceability, explainability, robustness, usability, privacy, continuous monitoring, and human oversight. The review proposes an AI Healthcare Transformation Framework in which data infrastructure and AI models generate decision intelligence that is embedded into clinical and operational workflows under governance and organizational-readiness controls. Operational resilience emerges when these capabilities enable anticipation, absorption, adaptation, recovery, and organizational learning. The review concludes that the strategic value of AI depends less on stand-alone algorithmic accuracy than on socio-technical integration, trustworthy governance, interoperable data, and measured improvements in patient and system outcomes.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

1. INTRODUCTION

Healthcare systems are being reshaped by a convergence of demographic pressure, workforce shortages, rising costs, chronic disease burden, growing patient expectations, and a rapidly expanding digital data environment. Electronic health records (EHRs), imaging archives, laboratory systems, wearable sensors, genomic platforms, patient-generated data, and administrative systems now create a continuous stream of structured and unstructured information. The central managerial and clinical challenge is no longer simply obtaining data; it is converting heterogeneous data into timely, reliable, and actionable decisions. Artificial intelligence (AI) and health analytics have therefore become prominent components of digital-health strategy because they can support pattern recognition, forecasting, classification, language understanding, image interpretation, optimization, and decision support at scales that are difficult to achieve through manual analysis alone (Rajkomar et al., 2019; Topol, 2019).

AI in healthcare is an umbrella term that includes machine learning (ML), deep learning (DL), natural language processing (NLP), computer vision, large language models (LLMs), reinforcement learning, and related computational approaches. In clinical care, these methods are used for risk prediction, diagnosis, prognosis, treatment selection, image interpretation, and early-warning systems. At the organizational level, they are increasingly applied to admission prediction, length-of-stay estimation, bed and discharge forecasting, staffing, documentation, scheduling, and patient-flow management. When connected to reliable data and embedded within workflows, AI can become part of a learning infrastructure that continuously converts operational and clinical signals into decisions and feedback (Esteva et al., 2019; Secinaro et al., 2021; Spear et al., 2023).

Digitization converts analog information into digital form; digitalization uses digital technologies to improve a process; digital transformation changes how an organization creates value, coordinates work, makes decisions, and interacts with patients. AI can contribute to transformation when it changes the timing and quality of decisions, redistributes cognitive work, creates new forms of surveillance and prediction, and links previously fragmented data sources. Davenport and Glaser (2022) argue that adoption in provider organizations depends on distinctive regulatory, financial, clinical, and operational conditions, illustrating why technically successful models may diffuse slowly. Similarly, implementation studies in nursing and clinical operations show that perceived usefulness, workflow fit, trust, training, data availability, and role clarity shape adoption as much as algorithmic performance (Seibert et al., 2023; Sahoo et al., 2025).

The concept of decision intelligence provides a useful bridge between AI models and organizational outcomes. Decision intelligence can be defined as the systematic design of decision processes that combine data, models, domain knowledge, human judgment, and feedback. In healthcare, a decision-intelligence system does more than generate a risk score. It links prediction to an interpretable recommendation, presents the recommendation at the appropriate point in workflow, assigns responsibility for action, monitors the consequences of the decision, and learns from new outcomes. This framing addresses a persistent weakness in healthcare AI research: many studies optimize discrimination metrics such as area under the receiver-operating-characteristic curve (AUC) without demonstrating whether the model changes clinical behavior, patient outcomes, resource utilization, equity, or cost. Vollmer et al. (2020) therefore emphasized transparency, replicability, ethics, and effectiveness as critical questions for patient-benefiting AI.

Operational resilience is the second organizing concept of this review. A resilient healthcare organization can anticipate disruptions, absorb shocks while maintaining essential services, adapt resources and workflows, recover performance, and learn from experience. The COVID-19 pandemic exposed weaknesses in capacity planning, supply chains, data interoperability, workforce flexibility, and real-time surveillance, and it accelerated interest in AI-supported forecasting and bedside decision support (Yu et al., 2021). The same logic applies beyond pandemics. Emergency departments experience volatile demand; inpatient wards must balance admissions and discharges; critical-care units need early detection of deterioration; and health systems must respond to cyber incidents, seasonal surges, staff shortages, and supply constraints.

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

Predictive models can strengthen resilience if they provide sufficiently early, reliable, and actionable information. For example, machine-learning models have been used for emergency admission prediction, aggregate demand forecasting, and inpatient discharge forecasting, creating opportunities to convert local predictions into capacity-management actions (King et al., 2022; Gao et al., 2022; Tschoellitsch et al., 2023).

The benefits of AI are accompanied by significant risks. Clinical data are generated through care processes rather than controlled experiments, and they contain missingness, measurement error, coding practices, access disparities, and institutional patterns that may be learned by algorithms. Bias can therefore be introduced through problem formulation, data collection, labeling, model development, evaluation, deployment, and feedback. Celi et al. (2022) and Chen et al. (2023) describe how data and algorithmic choices can perpetuate existing health disparities. Explainability is also contested: post-hoc explanations may be useful for debugging or communication but do not automatically establish causal validity. Privacy and security concerns are amplified when models are trained across institutions or use sensitive longitudinal records. Federated learning has emerged as one strategy for collaborative model development without centralizing raw data, but it does not eliminate governance, security, or representativeness challenges (Lee et al., 2021; Cremonesi et al., 2023).

Generative AI and LLMs have added a new layer to this landscape. Transformer-based models can summarize notes, answer clinical questions, generate documentation, and act as natural-language interfaces to complex information systems. Studies and reviews from 2024-2026 report rapid expansion of these applications while highlighting hallucination, privacy, provenance, explainability, evaluation, and over-reliance as unresolved issues (Nerella et al., 2024; Wang et al., 2024; Bracken et al., 2025; Collaco et al., 2026). Recent governance frameworks accordingly shift the focus from model accuracy to lifecycle trustworthiness. The FUTURE-AI international consensus guideline emphasizes fairness, universality, traceability, usability, robustness, and explainability across design, validation, deployment, and monitoring (Lekadir et al., 2025).

Despite a large and growing literature, studies are often siloed by specialty or technology. Reviews may focus on diagnosis, imaging, nursing, explainable AI, LLMs, or ethics, while operational studies focus on specific hospital problems. Less attention has been given to how these strands jointly contribute to digital transformation, decision intelligence, and operational resilience. This review addresses that gap through an integrated synthesis. It asks four questions: (1) Which AI and health-analytics applications are most relevant to modern healthcare-system transformation? (2) How do these applications contribute to decision intelligence at clinical and organizational levels? (3) Through what mechanisms can AI strengthen or weaken operational resilience? and (4) Which technical, organizational, ethical, and governance conditions determine whether AI creates sustainable healthcare value?

2. Systematic Review Methodology

2.1 Review design and reporting approach

This study was conducted as a focused systematic review and thematic synthesis guided by PRISMA 2020 principles. The objective was not to estimate a single pooled diagnostic or prognostic effect; instead, the review integrated heterogeneous evidence across clinical prediction, EHR intelligence, imaging, operational analytics, implementation, explainability, governance, and emerging generative AI. Because the topic spans clinical medicine, biomedical informatics, health services research, operations, and technology governance, a broad evidence design was necessary. Empirical model-development and validation studies were included alongside high-quality systematic reviews, implementation studies, and consensus/guideline papers when they directly informed healthcare-system transformation or trustworthy deployment.

2.2 Information sources and search strategy

The final structured search was completed on August 7, 2026. PubMed/MEDLINE served as the primary biomedical index. Supplementary searches used publisher- and index-based discovery across ScienceDirect/Elsevier, SpringerLink, BMJ, and IEEE-indexed literature, followed by backward and forward citation chaining from high-relevance reviews and consensus papers. Search concepts were

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

grouped into four blocks: (A) AI methods ("artificial intelligence," "machine learning," "deep learning," "natural language processing," "large language model," "generative AI," "computer vision," "federated learning," "explainable AI"); (B) data and analytics ("health analytics," "electronic health record," "medical imaging," "predictive analytics," "clinical data"); (C) decision and transformation terms ("clinical decision support," "decision intelligence," "digital transformation," "learning health system"); and (D) operational terms ("patient flow," "triage," "admission," "readmission," "discharge," "capacity," "operational resilience"). Boolean combinations were adapted to platform syntax.

The review emphasized literature published from January 2020 through August 2026, but a limited set of foundational studies from 2018-2019 was retained when necessary to establish major methodological or conceptual milestones, including scalable EHR deep learning, autonomous diabetic-retinopathy screening, and foundational reviews of AI in medicine (Rajkomar et al., 2018; Abràmoff et al., 2018; Topol, 2019). This strategy reduced historical over-expansion while preserving continuity between early clinical AI and contemporary system-level deployment.

2.3 Eligibility criteria

Studies were eligible when they were peer reviewed; written in English; focused on human healthcare, health-service delivery, or healthcare-system operations; and examined an AI or advanced analytics method with relevance to prediction, diagnosis, documentation, decision support, workflow, operations, implementation, governance, or resilience. Empirical studies needed to report a defined dataset or clinical setting and a describable analytical method. Reviews and consensus papers were retained when their methods were explicit and their scope directly informed the review questions. Exclusion criteria were non-healthcare applications, purely technical algorithm papers without healthcare evaluation, editorials without substantive analytical contribution, abstracts without adequate methodological information, duplicate reports, and studies whose primary focus could not be linked to healthcare decisions or system performance.

2.4 Screening and study selection

The structured search and citation-chaining process yielded a curated corpus of 68 candidate records. Six duplicate records were removed. Titles and abstracts of 62 unique records were screened against the eligibility criteria, resulting in 17 exclusions for insufficient relevance to the integrated healthcare-system scope. Forty-five reports underwent full-text or complete-abstract eligibility assessment. Five were excluded at this stage: two were outside the health-system transformation focus, two provided insufficient methodological detail for meaningful synthesis, and one was a duplicate secondary report. Forty core studies were retained for the evidence matrix and thematic synthesis. Additional peer-reviewed papers that provided foundational concepts, methodological standards, or contextual evidence were used in the introduction and discussion but were not counted in the 40-study core matrix. Figure 1 presents the PRISMA-style flow.

2.5 Data extraction and quality appraisal

For each core study, the following fields were extracted: authors and year, clinical or organizational setting, dataset or evidence source, AI method, principal finding, and limitation relevant to external validity or implementation. For predictive-model studies, quality appraisal was informed by PROBAST domains, particularly participant selection, predictor measurement, outcome definition, and analytical validity. Reporting expectations from TRIPOD-AI and related clinical-AI guidance were considered when judging transparency. Diagnostic/imaging studies were reviewed for reference-standard clarity, spectrum and sampling concerns, and external validation. Systematic reviews were considered using AMSTAR 2 principles such as search transparency, eligibility specification, duplicate processes, and risk-of-bias consideration. Implementation and governance studies were appraised pragmatically using CASP/JBI-style criteria addressing setting, stakeholder sampling, transparency of methods, and transferability. Given the heterogeneity of designs, no single numerical quality score was calculated; instead, recurrent limitations were carried forward into the synthesis.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

2.6 Synthesis strategy

Athematic synthesis was used because effect measures, outcomes, and study designs were not sufficiently homogeneous for meta-analysis. Studies were first coded according to primary function: predictive clinical analytics; EHR/NLP/LLM intelligence; imaging and multimodal diagnostics; precision/population/privacy-preserving analytics; operational analytics; or governance/decision intelligence. A second coding pass identified cross-cutting implementation dimensions: data quality, interoperability, workflow integration, explainability, fairness, privacy, external validation, organizational readiness, and monitoring. The final analytical framework connected these themes to three higher-order constructs: digital transformation (changes in processes and value creation), decision intelligence (integration of models with decisions and feedback), and operational resilience (anticipation, absorption, adaptation, recovery, and learning).

Table 1. Search strategy and information sources

Source	Core concepts	Date range	Role in review
PubMed/MEDLINE	AI/ML/DL/NLP/LLM + EHR/imaging/analytics + clinical decision support/operations	2018-Aug 2026; emphasis 2020-2026	Primary biomedical search and verification of peer-reviewed records/DOIs
ScienceDirect / Elsevier indexed literature	Clinical prediction, imaging, EHR, XAI, digital twins, LLMs	2020-Aug 2026	Supplementary discovery and cross-checking
SpringerLink	Healthcare delivery, implementation, digital transformation, XAI	2020-Aug 2026	Supplementary implementation and systems literature
BMJ / medical journal publisher search	Clinical AI governance, pandemic response, reporting standards	2020-Aug 2026	Governance, patient benefit, trustworthy deployment
IEEE-indexed biomedical literature	Medical imaging, federated learning, forecasting, health informatics	2020-Aug 2026	Engineering and operational analytics evidence
Citation chaining	Reference lists and related-article trails from key reviews and consensus papers	No fixed start date; foundational studies retained selectively	Identify seminal studies and recent connected evidence

Table 2. Inclusion and exclusion criteria

Domain	Included	Excluded
Publication	Peer-reviewed journal articles; empirical studies, systematic reviews, implementation studies, consensus/guideline papers	Editorials without substantive analysis; conference abstracts with insufficient detail; duplicate reports
Time	Primarily 2020-2026; selected 2018-2019 foundational studies	Older literature not necessary for conceptual continuity
Population/setting	Human healthcare, health-service delivery,	Non-healthcare AI; purely industrial or technical benchmarks

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Domain	Included	Excluded
	hospital/clinic/population-health settings	
AI/analytics	ML, DL, NLP/LLM, computer vision, federated learning, XAI, predictive/operational analytics	No meaningful AI/advanced analytics component
Outcomes	Clinical prediction/diagnosis, workflow, decisions, operations, implementation, governance, resilience	Technical performance disconnected from healthcare application or decision context

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

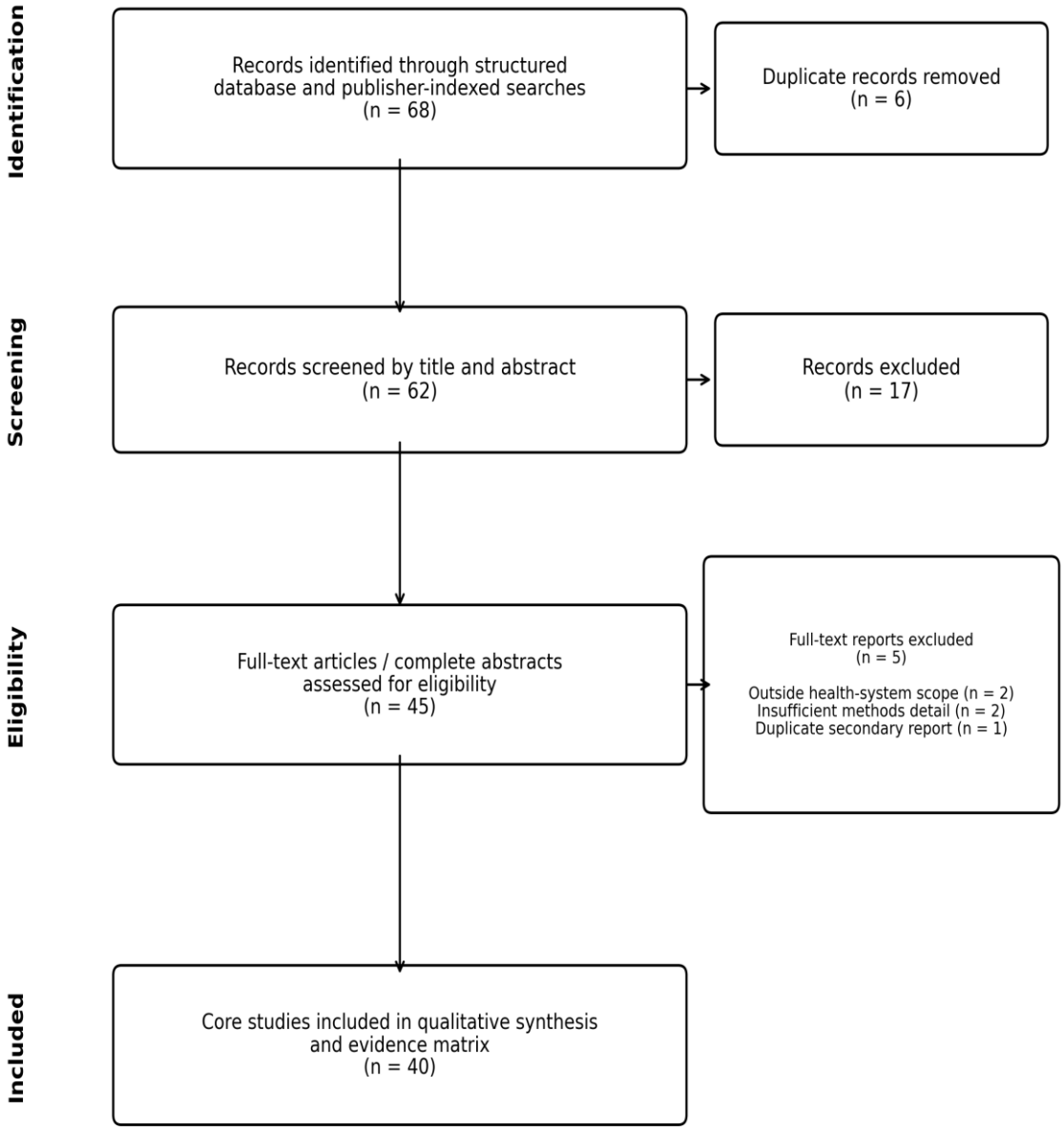


Figure 1. PRISMA-style flow diagram for the focused systematic review corpus.

3. Results and Thematic Synthesis

3.1 Characteristics of the evidence base

The 40-study core corpus was heterogeneous by design, reflecting the multi-level nature of healthcare AI. Empirical studies ranged from single- or multi-center retrospective EHR cohorts to population-scale cardiovascular datasets, emergency-department records, imaging studies, and federated-learning validations. Review and consensus studies addressed explainability, patient safety, fairness, implementation, and trustworthy deployment. The dominant methodological families were gradient-

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

boosting and tree ensembles, random forests, neural networks/deep learning, NLP/transformers, computer vision, federated learning, and structured consensus methods. Across the empirical literature, performance was frequently reported through AUC, sensitivity, specificity, accuracy, F1 score, calibration, or operational forecasting error. However, only a minority of studies directly measured downstream patient outcomes, clinician behavior, cost, or organization-level resilience. This gap is central to interpreting the evidence: AI capability is well documented, but transformation and resilience require evidence that predictions are translated into action.

3.2 AI-driven clinical and predictive analytics

Predictive analytics is one of the most mature healthcare AI domains because EHRs provide longitudinal observations that can be linked to clinically meaningful outcomes. Rajkomar et al. (2018) demonstrated that deep-learning approaches could use large-scale EHR data with relatively limited manual feature engineering to predict outcomes such as mortality, readmission, prolonged stay, and discharge diagnoses. This work helped establish the possibility of representation learning directly from complex clinical records. Subsequent studies have often favored tree-based methods for structured clinical data because they perform strongly on tabular variables and provide tractable feature-importance analyses.

Sepsis prediction illustrates both the promise and the limitations of the field. Wang et al. (2021) developed a random-forest model using 4,449 infected ICU patients and 55 EHR-derived features, reporting an AUC of 0.91, sensitivity of 87%, and specificity of 89% in the validation data. The authors explicitly called for external validation, highlighting a recurrent issue in clinical AI: high internal performance does not establish transportability. Moor et al. (2021), in a systematic review of early sepsis prediction, similarly found substantial methodological diversity and emphasized the need for standardized evaluation and clinically realistic prediction windows. Other sepsis studies extend the target from onset to mortality, delirium, and thrombocytopenia, showing how the same data infrastructure can support multiple risk-management decisions (Jiang et al., 2021; Jiang et al., 2022; Zhang et al., 2023; Alanazi et al., 2023).

Cardiovascular risk prediction provides a second example. Cho et al. (2021) evaluated conventional and ML models in 222,998 Korean adults and found that a neural-network model modestly improved discrimination over the pooled cohort equation. Xi et al. (2022) analyzed 143,043 Chinese patients with hypertension and reported improved prediction using ML compared with traditional logistic regression. Huang et al. (2022) incorporated lifestyle and wearable-derived information and found ensemble methods outperformed the Framingham Risk Score in their study population. At the same time, systematic comparative work indicates that performance gains over traditional risk models can be inconsistent and sensitive to design choices, outcome definitions, and validation strategy (Liu et al., 2023). These findings caution against equating algorithmic novelty with clinical superiority.

Readmission prediction demonstrates similar trade-offs. Lo et al. (2021) used ML to model 14-day unplanned readmission, while Davis et al. (2022) investigated machine-learned features from longitudinal administrative data. The practical importance of these models is clear: readmissions are costly and potentially preventable. Yet readmission is influenced by disease severity, discharge planning, social determinants, access to follow-up, and coding practices. A model may therefore identify vulnerability without identifying an effective intervention. From a decision-intelligence perspective, the value chain is incomplete unless risk estimates trigger an evidence-based action pathway and the health system measures whether the action changes outcomes.

Across predictive studies, three patterns emerge. First, ML can capture nonlinear interactions and high-dimensional structure that conventional scores may miss. Second, local performance is often strong enough to justify prospective evaluation. Third, many studies remain retrospective and setting-specific. Dataset shift is particularly important because clinical practice, laboratory reference ranges, coding, populations, and resource availability vary across institutions and over time. For operational deployment, calibration, external validation, and continuous monitoring may matter as much as peak discrimination. The recent FUTURE-AI guideline

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

formalizes this lifecycle view by emphasizing robustness, universality, traceability, and monitoring rather than one-time validation (Lekadir et al., 2025).

3.3 Electronic health records, natural-language intelligence, and generative AI

EHR intelligence has evolved from structured-feature prediction toward representation learning and natural-language interfaces. Traditional EHR models require harmonization of diagnoses, laboratory values, medications, procedures, and temporal sequences. Kiser et al. (2022) showed that standard vocabularies can improve model transferability for healthcare-associated infection prediction, underscoring the importance of semantic interoperability. Without common data models and consistent vocabularies, the same concept may be represented differently across hospitals, reducing portability and increasing implementation cost.

NLP addresses the large proportion of clinical information stored as free text. Clinical notes contain assessments, symptoms, reasoning, social context, and plans that may not be captured in structured fields. Transformer architectures can learn contextual representations of this text and support information extraction, classification, summarization, and question answering. Yang et al. (2022) described a large language model for EHRs, signaling a transition toward foundation-model approaches that can be adapted to multiple downstream tasks. Reviews of transformers and LLMs in healthcare emphasize potential applications in documentation, coding, patient communication, decision support, and knowledge retrieval while also identifying hallucination, bias, privacy, and evaluation as major challenges (Nerella et al., 2024).

The arrival of conversational generative AI has accelerated this transition. Wang et al. (2024) systematically reviewed applications and concerns of ChatGPT and related conversational LLMs in healthcare. The literature shows that LLMs can generate coherent explanations and answer examination-style questions, but performance in controlled knowledge tasks is not equivalent to safe clinical reasoning. Waldock et al. (2024) found substantial capability on healthcare examinations and certification-style tasks, yet such evaluations cannot capture patient-specific uncertainty, missing information, workflow constraints, or responsibility for errors. Bracken et al. (2025) reviewed AI-powered documentation systems and highlighted their potential to reduce documentation burden, a significant organizational concern because administrative workload contributes to clinician dissatisfaction and burnout.

From a digital-transformation perspective, the most important characteristic of LLMs may be interface rather than autonomous expertise. Kather et al. (2024) argued that large language models could restore natural language as a universal interface for healthcare information systems. If connected to validated clinical knowledge, EHR data, and retrieval systems, language models could lower the cognitive cost of navigating fragmented software. Retrieval-augmented generation is being investigated as a method for grounding model outputs in external evidence and reducing unsupported generation, although current systematic reviews still report heterogeneity in architecture, evaluation, and safety controls (Amugongo et al., 2025). Agentic AI further extends the concept by enabling models to plan and execute multi-step tasks across tools, but 2026 reviews emphasize that healthcare applications remain early and require rigorous oversight (Collaco et al., 2026; Njei et al., 2026).

The principal risk is that a fluent interface can make uncertainty less visible. Clinical users may over-trust plausible language, and patients may not distinguish a confident response from a validated one. Therefore, EHR/LLM intelligence should be designed around provenance, retrieval quality, role-based access, auditability, and escalation pathways. Human-AI collaboration is not a temporary compromise; it is a core safety architecture in domains where data are incomplete and consequences are high.

3.4 Medical imaging and multimodal diagnostic intelligence

Medical imaging is among the most visible success areas for deep learning because images are high-dimensional, standardized digital objects and diagnostic tasks often have clearly defined labels. Foundational studies demonstrated autonomous diabetic-retinopathy screening and large-scale deep-learning image classification, while later work expanded to pathology, radiology, ophthalmology, electroencephalography, and resource-limited

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

screening (Abràmoff et al., 2018; Esteva et al., 2019). Steiner et al. (2020) evaluated combined AI and pathologist assessment for prostate-biopsy review, illustrating a clinically important pattern: AI may create more value through augmentation than replacement. Holmström et al. (2021) demonstrated point-of-care digital cytology with AI for cervical-cancer screening in a resource-limited setting, showing how digital diagnostics can potentially extend specialist capacity.

Reviews of medical-image analysis show rapid advances in convolutional neural networks, transformers, segmentation, detection, and multimodal models, but also persistent external-validity concerns (Chan et al., 2020; Chen et al., 2022). Imaging datasets can encode scanner type, acquisition protocol, institution, disease prevalence, and referral patterns. Models may therefore exploit shortcuts that are associated with labels in development data but are not stable across settings. The problem becomes more important as models move from retrospective benchmark datasets to clinical deployment.

Tveit et al. (2023) evaluated automated interpretation of clinical electroencephalograms, indicating that AI-supported interpretation is expanding beyond conventional image modalities. Ophthalmology has similarly become a testbed for autonomous and semi-autonomous systems because optical coherence tomography and retinal imaging are highly digitized (Popescu Patoni et al., 2023). Clinical-evaluation guidelines increasingly emphasize prospective validation, representative populations, comparator selection, and real-world workflow testing (Yang et al., 2023).

Explainability in imaging is complex. Saliency maps and feature visualizations can indicate image regions associated with a prediction, but they may be unstable or provide false reassurance. Reviews by Ali et al. (2023), Jung et al. (2023), Loh et al. (2022), and Chaddad et al. (2023) conclude that explanation usefulness depends on the user, task, and decision context. A radiologist may need localization and uncertainty; a patient may need an understandable rationale; a regulator may need traceability and performance evidence. The 2026 imaging-focused interpretation of FUTURE-AI reinforces the need to treat explainability as one element within a broader trustworthy-AI lifecycle (Kondylakis et al., 2026).

The next stage of imaging intelligence is multimodal. A diagnostic model may combine images with pathology, laboratory values, clinical notes, genomics, and longitudinal history. This could reduce false positives and produce more patient-specific recommendations, but it also increases data-integration complexity and makes error analysis harder. Multimodal systems therefore depend on interoperability and robust data governance to a greater degree than single-modality classifiers.

3.5 Precision, population, and privacy-preserving analytics

AI can support personalization by modeling heterogeneity across patients and combining clinical, behavioral, genomic, and sensor data. Precision-health applications range from treatment-response prediction to digital twins that simulate individualized trajectories. Digital-twin reviews describe a conceptual shift from static patient profiles to computational representations that are continuously updated as new data arrive (Armeni et al., 2022; Sun et al., 2023; Machado & Berssaneti, 2023). Coorey et al. (2022) discuss health digital twins in cardiovascular disease, where physiological models and data-driven learning could eventually support personalized risk estimation and intervention simulation. However, digital twins remain an emerging field with major validation and governance requirements; a model that simulates an individual creates a strong appearance of precision even when uncertainty is substantial.

Population-health analytics extends the unit of analysis from individual patients to communities and health systems. AI can be used for disease surveillance, resource targeting, risk stratification, and identification of unmet need. The ethical stakes are correspondingly broader because population models can influence access, prioritization, and public-health interventions. Couture et al. (2023) found that governance of AI in population health should include public perspectives, not only technical and clinical experts. Fairness cannot be reduced to a single statistical metric because priorities may conflict: equalized performance, equal access, equal benefit, and protection of vulnerable groups are not always simultaneously achievable.

Privacy-preserving learning is particularly relevant when high-quality models require data from multiple

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

institutions. Federated learning allows local sites to train shared models without transferring raw data to a central repository. Lee et al. (2021) validated a federated approach for thyroid ultrasound analysis in a real healthcare environment, demonstrating the practical feasibility of collaborative image learning while protecting local data. Almufareh et al. (2023) applied federated learning to breast-cancer prediction, and Darzidehkalani et al. (2022) described its potential for multicenter medical-imaging ecosystems. Cremonesi et al. (2023) further argued for multimodal federated healthcare platforms.

Federated learning should not be treated as a complete privacy solution. Model updates can leak information under some conditions, participating institutions can have highly non-identically distributed data, and governance is still required for consent, accountability, security, model ownership, and benefit sharing. Privacy is therefore one design constraint within a broader socio-technical system. Secure aggregation, differential privacy, access controls, auditing, and legal agreements may be needed depending on the use case.

3.6 Hospital operational analytics and operational resilience

Operational analytics is where AI most directly connects to health-system resilience. Emergency departments, inpatient wards, operating rooms, diagnostic services, and staffing systems are coupled through shared resources. Delays in one part of the system can propagate elsewhere. Traditional operations research has long addressed queuing, scheduling, and capacity, but ML adds the ability to learn demand and patient-level risk from complex data. The combination can support anticipatory management rather than reactive firefighting.

Emergency-department admission prediction is a prominent application. De Hond et al. (2021) compared ML and conventional approaches for hospital-admission prediction across multiple hospitals, demonstrating that early information can support disposition planning. Nguyen et al. (2021) developed models to personalize care levels among emergency patients, while Patel et al. (2022) used an inclusive gradient-boosting model that incorporated clinical variables and free-text triage information. Yu et al. (2022) externally validated an interpretable ML-based emergency risk score, an important contribution because external validation is uncommon in operational AI. Tschoellitsch et al. (2023) used triage information for admission and mortality prediction, and Chang et al. (2022) explored identification of low-severity patients likely to have short discharge stays.

Operational value increases when predictions are aggregated to the system level. King et al. (2022) demonstrated real-time aggregated prediction of hospital admission for emergency patients, which is directly relevant to bed management and staffing. Gao et al. (2022) used ML to forecast inpatient discharges for Singapore hospitals, supporting estimation of future bed availability. Reviews of machine learning in patient flow show applications across demand prediction, transfers, length of stay, discharge timing, and resource planning, but they also note the absence of shared benchmark datasets and limited evidence that predictive improvements translate to system-wide gains (El-Bouri et al., 2021).

These studies support a resilience mechanism with five stages. First, AI can improve anticipation by forecasting demand, deterioration, or resource need. Second, it can support absorption by prioritizing scarce capacity during surges. Third, it can facilitate adaptation by updating schedules, routing, or staffing as conditions change. Fourth, it can support recovery through post-event analysis and prioritization of backlogs. Fifth, repeated monitoring can generate organizational learning. However, the same systems can reduce resilience if health organizations become over-dependent on fragile models, fail to maintain manual fallback processes, or ignore distribution shifts during crises.

Resilience therefore requires model monitoring and operational governance. A model trained during normal operations may perform poorly during a pandemic, cyber incident, policy change, or unusual surge because the relationship between predictors and outcomes changes. Alert fatigue is another risk: more predictions do not necessarily create better decisions. Operational AI should be evaluated through end-to-end metrics such as waiting time, length of stay, cancellation rates, boarding time, staff workload, safety events, and equity, not only AUC.

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

3.7 Decision intelligence, governance, and human-AI collaboration

The synthesis indicates that healthcare AI creates value when prediction is embedded in a governed decision process. Decision intelligence can be conceptualized as a pipeline: data acquisition -> model inference -> explanation and uncertainty -> recommendation -> human or automated action -> outcome measurement -> feedback and monitoring. Each transition can fail. High-quality data can feed an inappropriate model; an accurate model can present an unusable interface; a clear recommendation can arrive too late; a clinician can correctly reject a model output but leave no feedback trace; or an organization can deploy a model without monitoring drift.

Governance is therefore not an administrative add-on. Morley et al. (2022) described international governance needs for healthcare data and AI, and Liao et al. (2022) reported an operational governance structure within a large health system spanning validity, acceptability, deployment, and continuous monitoring. Zhang and Zhang (2023) organize trustworthy medical AI around data quality, bias, opacity, safety/security, and responsibility. Rogers et al. (2021) show why case-based ethical analysis can reveal patient-care issues not captured by performance statistics.

Fairness is particularly important because health systems already contain structural inequities. Chen et al. (2023) provide a detailed account of algorithmic fairness in medicine, while Celi et al. (2022) emphasize global sources of bias. A model may be statistically accurate overall while systematically underperforming for a minority group. Bias can also arise from using cost or utilization as a proxy for health need, because access to care is itself unequal. Governance should therefore require subgroup evaluation, dataset documentation, monitoring for disparate impact, and mechanisms for contestability.

Explainability contributes to trust but should be aligned with decision needs. Ali et al. (2023) and Jung et al. (2023) show that explanation effectiveness is rarely measured consistently. Clinical users need concise, calibrated explanations that indicate important factors, uncertainty, and actionable alternatives rather than technically impressive visualizations. Human factors matter: too much information can increase cognitive load, while oversimplified explanations can encourage automation bias.

Recent consensus work provides a more mature deployment model. FUTURE-AI identifies fairness, universality, traceability, usability, robustness, and explainability as core principles and links them to practical recommendations across the AI lifecycle (Lekadir et al., 2025). Trust research likewise emphasizes that technical reliability, institutional accountability, transparency, and stakeholder relationships interact (Starke et al., 2025). The implication for decision intelligence is clear: trustworthy AI is an organizational capability that combines model engineering, clinical governance, ethics, cybersecurity, human factors, and performance management.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 3. Comparison of major AI healthcare application domains

Domain	Typical data	AI methods	Primary decisions	Potential value	Key risks
Predictive clinical analytics	EHR, labs, vitals, history	RF, gradient boosting, neural networks	Deterioration, sepsis, readmission, CVD risk	Earlier intervention and risk stratification	Dataset shift, calibration, alert burden
EHR/NLP/LLM intelligence	Notes, structured EHR, orders	Transformers, LLMs, NLP	Summarization, documentation, coding, information retrieval	Reduced information friction and documentation burden	Hallucination, privacy, provenance, over-trust
Imaging diagnostics	X-ray, CT, MRI, pathology, retinal/EEG signals	CNNs, vision transformers, multimodal models	Detection, grading, prioritization	Specialist augmentation and scalable screening	Shortcut learning, scanner/site bias, explainability
Precision/population analytics	Genomics, sensors, claims, population data	Ensembles, multimodal ML, digital twins	Personalization, surveillance, targeting	Tailored intervention and population planning	Equity, privacy, uncertain causal validity
Federated/privacy-preserving AI	Distributed institutional datasets	Federated learning, secure aggregation	Collaborative model training	Scale without raw-data centralization	Non-IID data, security, governance complexity
Operational analytics	Triage, admissions, discharges, capacity, staffing	Gradient boosting, time-series ML, optimization	Patient flow, beds, staffing, surge planning	Efficiency and operational resilience	Workflow dependency, drift during disruptions

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

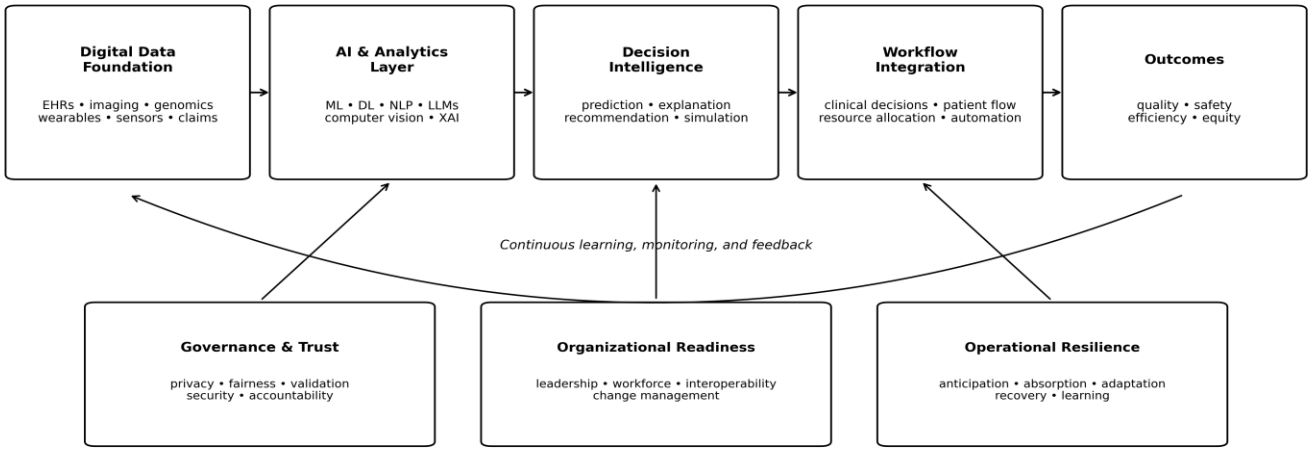


Figure 2. AI Healthcare Transformation Framework linking data, AI, decision intelligence, workflow integration, governance, organizational readiness, and operational resilience.

Table 4. Evidence matrix of 40 reviewed studies

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Rajkomar et al.	2018	Large multi-hospital EHR corpus	Deep neural networks / representation learning	Predicted mortality, readmission, length of stay and diagnoses directly from EHR data; demonstrated scalable end-to-end learning.	Retrospective; portability and interpretability require local validation.	10.1038/s41746-018-0029-1
Abràmoff et al.	2018	Primary-care retinal images	Autonomous deep-learning diagnostic system	Pivotal clinical trial supported autonomous detection of diabetic retinopathy in primary care.	Task-specific; performance depends on image quality and defined referral pathway.	10.1038/s41746-018-0040-6
Steiner et al.	2020	Prostate biopsy pathology slides	Deep learning + pathologist assessment	AI-assisted review improved efficiency/diagnostic support for prostate biopsy grading.	Specialist pathology setting; generalization depends on slide preparation and workflow.	10.1001/jamanetworkopen.2020.23267

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Holmström et al.	2021	Point-of-care cervical cytology images in a resource-limited setting	Computer vision / deep learning	Demonstrated AI-supported digital cytology for cervical cancer screening at point of care.	Deployment depends on image acquisition quality and local screening pathways.	10.1001/jamanetworkopen.2021.1740
Cho et al.	2021	222,998 Korean adults	Neural network, RF, AdaBoost, tree bagging, logistic regression	Neural network modestly improved cardiovascular risk discrimination over conventional risk scoring.	Single national population; benefit over conventional models was modest.	10.1038/s41598-021-88257-w
Wang et al.	2021	4,449 infected ICU patients; 55 EHR features	Random forest	Sepsis prediction achieved AUC 0.91, sensitivity 87%, specificity 89%.	Single-center retrospective cohort; external validation needed.	10.3389/fpubh.2021.754348
Jiang et al.	2021	Sepsis survivors with ICU readmission	Explainable machine learning	Identified risk factors for in-hospital mortality and provided interpretable risk analysis.	Retrospective ICU data; causal interpretation of feature importance is limited.	10.1016/j.cmpb.2021.106040
De Hond et al.	2021	Three hospitals; Netherlands Emergency Department Evaluation Database	Multiple ML models vs regression	Showed feasibility of early ED hospitalization prediction and compared ML with conventional approaches.	Performance depends on data available at each time point and local ED practice.	10.1016/j.ijmedinf.2021.104496

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Nguyen et al.	2021	Emergency-department patient records	Supervised ML for care-level prediction	Developed models to personalize care levels and support admission decisions.	Retrospective evaluation; prospective workflow impact not established.	10.1093/jamia/ocab118
Lee et al.	2021	Multi-institution thyroid ultrasound images	Federated deep learning	Validated federated image analysis in a real healthcare environment without centralizing raw images.	Institutional heterogeneity and federation infrastructure remain challenges.	10.2196/25869
Lo et al.	2021	Hospital discharge records	Machine-learning classification	Predicted 14-day unplanned readmission and identified higher-risk discharges.	Readmission is influenced by social and post-discharge factors not fully captured in EHR data.	10.1186/s12911-021-01639-y
Huang et al.	2022	600 Southeast Asian participants; lifestyle, wearables and calcium score	Ensemble ML	Ensemble models outperformed Framingham Risk Score for study-defined risk categories.	Relatively small cohort; external validation and calibration needed.	10.1038/s41598-021-04649-y
Xi et al.	2022	143,043 Chinese patients with hypertension	Multiple ML algorithms vs logistic regression	ML outperformed traditional logistic regression for current cardiovascular disease risk prediction.	Population-specific and cross-sectional/current-risk framing limits transportability.	10.3389/fcvm.2022.1025705

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Ordikhani et al.	2022	Cardiovascular risk cohort	Evolutionary machine learning	Developed an evolutionary ML approach for cardiovascular risk prediction.	Requires broader external validation and comparison under common clinical utility metrics.	10.1371/journal.pone.0271723
Davis et al.	2022	Population-level linked administrative hospital data, Alberta	Machine-learned longitudinal features + predictive models	Automatically learned longitudinal features improved readmission prediction workflow.	Administrative data may omit clinical and social variables; geographic transfer uncertain.	10.1186/s12913-022-08748-y
Chang et al.	2022	Emergency department triage data	Machine-learning classification	Identified low-severity ED patients likely to have short discharge length of stay, supporting streaming.	Local triage practices may limit generalizability.	10.1186/s12873-022-00632-6
Yu et al.	2022	External emergency-department validation cohort	Interpretable ML risk score (SERP)	Externally validated an interpretable triage score for emergency risk prediction.	External validation was setting-specific; impact on decisions/outcomes not directly tested.	10.1038/s41598-022-22233-w
Patel et al.	2022	Mount Sinai ED data, 2015-2019; tabular features + triage text	Gradient boosting + text features	Accurately predicted adult hospital admission at multiple ED time points.	Retrospective health-system data; deployment benefit requires prospective testing.	10.3390/jcm11236888

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
King et al.	2022	Real-time emergency patient data	Machine learning + aggregate forecasting	Produced real-time aggregate hospital-admission forecasts relevant to operational capacity planning.	Operational gains depend on integration with bed-management decisions.	10.1038/s41746-022-00649-y
Gao et al.	2022	Singapore hospital discharge time series	Machine-learning forecasting	Improved inpatient discharge forecasting for resource and bed-capacity planning.	Forecasting is sensitive to policy and demand shifts.	10.1109/JBHI.2022.3172956
Kiser et al.	2022	EHR data for healthcare-associated infection	ML with standardized vocabularies	Standard vocabularies improved transferability of ML models across EHR contexts.	Semantic standardization does not remove institutional distribution shift.	10.2196/39057
Yang et al.	2022	Longitudinal electronic health records	Large language model / transformer	Developed an EHR-focused large language model supporting multiple downstream health tasks.	Foundation-model outputs require task-specific validation, privacy controls and calibration.	10.1038/s41746-022-00742-2
Darzidehkalani et al.	2022	Multicenter medical-imaging ecosystem literature	Federated learning review	Outlined federated learning as a pathway toward multicentral imaging AI without centralizing raw data.	Review evidence was early and implementation standards remained immature.	10.1016/j.jacr.2022.03.015

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Tschoellitsch et al.	2023	Emergency-department triage records	Machine-learning prediction	Used triage data for admission and mortality prediction, supporting earlier resource decisions.	Performance depends on triage data quality and local workflows.	10.1097/MEJ.0000000000001068
Vântu et al.	2023	Emergency-department triage data	Supervised machine learning	Demonstrated ML processing of triage information for emergency-service decision support.	Limited evidence of prospective clinical impact.	10.1016/j.heliyon.2023.e18402
Alanazi et al.	2023	ICU sepsis data / reviewed ML prediction setting	Machine-learning sepsis prediction	Showed the continued value of ML for earlier sepsis identification in ICU settings.	Model comparability and external validation remain concerns.	10.3390/medicina59071276
Zhang et al.	2023	ICU sepsis cohort	Machine-learning risk prediction	Developed a model for sepsis-associated delirium risk to support targeted monitoring.	Retrospective design; delirium definitions and local practice can vary.	10.1038/s41598-023-38650-4
Almufareh et al.	2023	Breast-cancer prediction data	Federated learning	Demonstrated collaborative breast-cancer prediction under a federated-learning framework.	Federated experimental setting; real multi-institution operational complexity remains.	10.3390/healthcare11243185

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Tveit et al.	2023	Clinical electroencephalograms	Deep learning / automated interpretation	Evaluated AI-based automated interpretation of clinical EEGs, extending AI diagnostics beyond imaging.	Specialist-data quality and generalizability require continued prospective evaluation.	10.1001/jamaneurol.2023.1645
Cremonesi et al.	2023	Multimodal health data platform	Federated multimodal learning	Proposed and evaluated practical multimodal modeling for a federated healthcare platform.	Complexity of harmonizing multimodal, multi-site data remains high.	10.1016/j.jbi.2023.104338
Ali et al.	2023	Systematic literature review of XAI in healthcare	Systematic review	Synthesized XAI methods and emphasized explainability for trust and high-stakes decisions.	Heterogeneous explanation methods and limited user-centered evaluation.	10.1016/j.compbiomed.2023.107555
Jung et al.	2023	Systematic review of healthcare XAI	Systematic review	Identified properties needed for effective healthcare explanations and gaps in evaluation.	Explanation effectiveness measures were inconsistent across studies.	10.1016/j.heliyon.2023.e16110
Chen et al.	2023	Medicine and healthcare fairness literature	Critical review / framework	Detailed sources, definitions and mitigation strategies for algorithmic fairness in healthcare AI.	Fairness criteria can conflict; technical mitigation cannot resolve structural inequity alone.	10.1038/s41551-023-01056-8

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Spear et al.	2023	Health-care delivery applications	Review / health-system analysis	Mapped AI applications across clinical and delivery workflows and highlighted implementation value.	Evidence of operational and economic impact remains uneven.	10.1007/s10916-023-02018-y
Nerella et al.	2024	Transformer and LLM healthcare literature	Narrative/systematic review of transformers and LLMs	Synthesized LLM applications, architectures and risks across healthcare.	Rapidly evolving field; benchmarks and safety evaluation lack standardization.	10.1016/j.artmed.2024.102900
De Micco et al.	2024	Patient-safety AI literature	Systematic review	Identified potential for AI to support patient safety while emphasizing governance and human oversight.	Heterogeneous evidence and limited real-world safety endpoints.	10.3389/fmed.2024.1522554
Wang et al.	2024	Conversational LLM healthcare studies	Systematic review	Mapped ChatGPT/LLM applications and concerns including accuracy, privacy, bias and misuse.	Fast-moving evidence base and heterogeneous evaluation tasks.	10.2196/22769
Lekadir et al.	2025	117 experts from 50 countries; consensus process	International consensus guideline	Defined FUTURE-AI principles and 30 lifecycle practices for trustworthy and deployable healthcare AI.	Consensus guidance requires local operationalization and measurable implementation indicators.	10.1136/bmj-2024-081554

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Study	Year	Dataset / setting	AI method	Main finding	Key limitation	DOI
Lanfranchi et al.	2026	Systematic review of AI adoption in health services	Systematic review	Synthesized implementation barriers, enablers and health-services implications of AI adoption.	Implementation evidence remains heterogeneous across organizations and technologies.	10.1186/s12913-026-14960-x
Collaco et al.	2026	Agentic AI healthcare literature	Scoping review	Mapped emerging agentic AI use cases and highlighted governance and safety needs for autonomous workflows.	Evidence is early; few mature real-world clinical deployments.	10.1038/s41746-026-02517-5

4. Discussion

4.1 From algorithmic performance to digital transformation

The main finding of this review is that AI-enabled healthcare transformation is best understood as a socio-technical reconfiguration of decisions rather than a collection of high-performing models. The literature provides strong evidence that ML and DL can predict many clinically and operationally relevant outcomes. Yet the translation gap remains substantial. A model can improve AUC without improving health-system performance if it duplicates existing knowledge, produces non-actionable alerts, arrives at the wrong point in workflow, or lacks a clear response protocol. Digital transformation occurs when organizations redesign processes around reliable information, clarify responsibility, and measure downstream effects.

This distinction helps explain why AI adoption in provider organizations has been slower than technological progress might suggest. Healthcare organizations must integrate AI with legacy information systems, regulatory requirements, professional norms, reimbursement structures, procurement, data governance, and clinical accountability. Davenport and Glaser (2022) identify these institutional conditions as major determinants of adoption. The 2026 systematic review by Lanfranchi et al. similarly emphasizes implementation challenges and health-services implications, suggesting that organizational barriers remain central even as algorithms improve.

The practical unit of transformation should therefore be the decision pathway. For an emergency-admission model, the relevant pathway includes triage data capture, model inference, bed-management visibility, escalation thresholds, clinician review, and capacity actions. For an imaging model, it includes acquisition, quality control, model output, reader interaction, discrepancy management, reporting, and audit. For an LLM documentation tool, it includes audio or text capture, draft generation, verification, correction, signature, storage, and monitoring of errors. Mapping the full pathway makes hidden dependencies visible and creates a basis for evaluating whether AI reduces burden or merely shifts work.

4.2 Decision intelligence as the operating model for healthcare AI

Decision intelligence offers a unifying operating model because it connects analytics to action and learning. Traditional clinical decision-support systems often use rules or thresholds. Modern AI expands the range of signals that can be processed, but it also increases opacity and complexity. A

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

mature decision-intelligence architecture should therefore include five features: calibrated prediction, meaningful explanation, decision ownership, workflow integration, and feedback. These features help prevent two extremes: blind automation and symbolic AI adoption with little operational effect.

Human-AI collaboration should be designed according to task risk and uncertainty. Low-risk repetitive administrative tasks may tolerate greater automation, provided privacy and quality controls are in place. High-risk diagnosis and treatment decisions should generally preserve meaningful human oversight. The appropriate level of autonomy can change as evidence accumulates. Autonomous diabetic-retinopathy screening demonstrates that carefully bounded, validated tasks can achieve substantial autonomy, whereas open-ended LLM reasoning remains difficult to constrain. The key is not whether humans are "in the loop" in name, but whether they possess the information, authority, time, and competence to detect and correct errors.

Decision intelligence also creates a mechanism for organizational learning. Every accepted, rejected, or overridden recommendation can become feedback if captured appropriately. Over time, health systems can identify drift, recurring false positives, subgroup differences, and workflow bottlenecks. This creates a learning health-system cycle in which data inform decisions, decisions generate outcomes, and outcomes update both models and processes.

4.3 Operational resilience: potential benefits and new dependencies

The evidence supports a credible pathway from AI to operational resilience, especially through forecasting and early warning. Admission and discharge predictions can give managers more time to adjust beds and staffing; sepsis and deterioration models can support earlier escalation; population analytics can identify emerging demand; and digital twins may eventually support scenario testing. During disruptions, the value of lead time can be substantial.

However, resilience is not the same as efficiency. Highly optimized systems can become brittle if they eliminate slack or depend on a narrow data pipeline. AI can create new single points of failure through cloud services, vendor APIs, centralized identity systems, or model dependencies. Cybersecurity incidents can disrupt data availability; a sudden workflow change can invalidate predictions; and staff may lose confidence after visible failures. Operational resilience therefore requires graceful degradation: clear fallback processes, local decision authority, downtime procedures, and the ability to operate safely when AI is unavailable.

Another concern is automation bias during crisis conditions. Under high workload, users may be more likely to accept AI recommendations without scrutiny. Conversely, frequent false alerts can create distrust and override. Resilient implementation should monitor both model performance and human response, including override rates, time to action, alert burden, and near misses.

4.4 Trustworthy governance, fairness, and accountability

The governance literature has matured considerably since early ethical discussions. Current frameworks emphasize lifecycle controls rather than one-time ethical approval. FUTURE-AI is particularly useful because its six principles can be translated into measurable practices. Fairness requires representative development and subgroup evaluation; universality addresses performance across settings; traceability requires documentation and audit; usability focuses on human factors and workflow; robustness addresses perturbation, drift, and failure; and explainability concerns understandable reasoning and communication (Lekadir et al., 2025).

These principles are interdependent. For example, an explainable model can still be unfair; a robust model can still violate privacy; and a well-calibrated model can be unusable. Governance structures should therefore include multidisciplinary review involving clinicians, data scientists, informaticians, operations leaders, cybersecurity, legal and ethics expertise, and patient/public representation where appropriate. High-impact systems should have predefined monitoring thresholds and retirement criteria.

Accountability deserves particular attention with generative and agentic AI. If an LLM drafts a note that a clinician signs, responsibility is legally and ethically shared in complex ways among the clinician, organization, vendor, and system designers. Agentic systems that call tools or execute actions increase the importance of permission boundaries, logging,

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

verification, and rollback. Recent 2025-2026 reviews show that enthusiasm for generative AI is increasingly matched by attention to governance, user trust, and patient acceptance (Henzler et al., 2025; Shi et al., 2026; Collaco et al., 2026).

4.5 Organizational readiness and workforce implications

Successful AI adoption requires organizational capabilities that are unevenly distributed across health systems. Data engineering, clinical informatics, machine-learning operations, cybersecurity, product management, human-factors design, and evaluation expertise are all needed. Smaller organizations may lack these capabilities and become dependent on vendors, increasing the importance of procurement standards and transparent performance claims.

Workforce readiness is equally important. Clinicians need enough AI literacy to interpret outputs, understand uncertainty, recognize potential bias, and use escalation pathways. They do not need to become data scientists, but they need task-specific competence. Managers need to understand that implementation requires workflow redesign and ongoing maintenance rather than one-time software installation. Health-professional perspective reviews consistently identify training, trust, perceived usefulness, workflow fit, and governance as facilitators or barriers (Sahoo et al., 2025; Henzler et al., 2025).

AI may also redistribute work. Documentation systems could reduce clerical burden but create new verification tasks. Imaging AI may accelerate routine cases while concentrating human attention on ambiguous cases. Predictive models may create new care-management queues. These changes should be measured through workload, cognitive burden, professional autonomy, and patient interaction, not only throughput.

4.6 Data interoperability and learning infrastructure

The review also shows that data architecture is a strategic constraint. Models trained on local coding and data structures are expensive to transfer. Standard vocabularies, common data models, and interoperability standards such as FHIR can reduce integration friction, but semantic consistency remains challenging. Kiser et al. (2022) demonstrate the value of standard vocabularies for transferability, while federated-learning research shows how collaboration can occur without centralizing raw data. A future learning infrastructure will likely combine interoperable local data, privacy-preserving collaboration, continuous monitoring, and model registries.

This infrastructure should support provenance. Every prediction should be traceable to model version, input data, time, and context. This is particularly important when models are updated frequently. Without version control and audit trails, organizations cannot determine which model influenced a decision or investigate harm. Traceability is therefore both a technical and governance requirement.

Table 5. Cross-cutting implementation challenges and mitigation priorities

Challenge	Why it matters	Mitigation priority	Evaluation indicator
Dataset shift and weak external validity	Performance can degrade across sites, time and populations	External/temporal validation; recalibration; drift monitoring	Calibration, subgroup performance, post-deployment drift
Algorithmic bias	Unequal errors can reinforce health disparities	Representative data; subgroup audit; fairness review; contestability	Error and benefit distribution by protected/clinically relevant groups
Opacity and poor explanations	Users may over-trust or under-use model outputs	Task-specific XAI; uncertainty display; human-factors testing	Comprehension, appropriate reliance, override quality
Privacy and cybersecurity	Health data and model interfaces are high-value targets	Data minimization; federated approaches;	Privacy incidents, access anomalies, security testing outcomes

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Challenge	Why it matters	Mitigation priority	Evaluation indicator
		access controls; threat modeling	
Interoperability	Local coding and fragmented systems limit transfer	FHIR/common models; standard vocabularies; semantic governance	Integration effort, data completeness, portability across sites
Workflow misfit and alert burden	Accurate models can fail if not actionable	Co-design; decision-pathway mapping; threshold optimization	Time-to-action, alert volume, acceptance/override rate
Accountability and lifecycle governance	Responsibility can be unclear after deployment or update	Multidisciplinary AI governance; model registry; audit trails	Documented owner, version traceability, monitoring/retirement criteria
Workforce readiness	Users need competence to interpret and challenge AI	Role-based AI literacy and simulation-based training	Knowledge, confidence, safe use, escalation behavior

5. Future Research Agenda

Future research should move from retrospective model comparison toward prospective, system-level evaluation. First, clinical AI studies should prioritize external and temporal validation, calibration, subgroup performance, and decision-curve or utility analysis. A small improvement in AUC may have little practical value if false positives create excessive workload. Second, more studies should use pragmatic trials, stepped-wedge designs, interrupted time series, or carefully designed prospective observational studies to test whether AI changes care processes and outcomes.

Third, economic evaluation should become routine. The 2026 systematic review of economic evidence in healthcare AI indicates that evidence on cost-effectiveness remains less mature than evidence on technical performance (Schmidt et al., 2026). Health systems need information about implementation costs, infrastructure, training, maintenance, licensing, error costs, and opportunity costs. Fourth, operational-resilience research should evaluate AI under stress conditions. Models should be tested during demand surges, workflow changes, missing data, cyber downtime, and population shifts rather than only under stable retrospective conditions.

Fifth, generative and agentic AI require new evaluation paradigms. Static benchmark accuracy is insufficient for systems that generate free text or execute multi-step actions. Evaluation should include factuality, provenance, uncertainty communication, tool-use safety, permission compliance, recovery from errors, adversarial robustness, and human factors. Retrieval-augmented systems should report retrieval quality separately from generation quality. Agentic systems should have bounded autonomy, audit logs, and sandboxed permissions before clinical deployment.

Sixth, research should explicitly examine equity and global applicability. Models developed in high-resource academic centers may not transfer to rural, lower-resource, or linguistically diverse settings. Federated learning and open standards may facilitate collaboration, but they must be paired with local validation and governance. Finally, digital twins should be evaluated cautiously. Their promise lies in integrating mechanistic and data-driven modeling for individualized simulation, but clinical claims should match validation evidence. The field should distinguish exploratory simulation from decision-grade prediction.

The proposed AI Healthcare Transformation Framework (Figure 2) summarizes these priorities. Digital data provide the foundation; AI and analytics generate predictions and representations; decision intelligence converts them into recommendations; workflow integration turns recommendations into action; and outcomes feed continuous learning. Governance and trust, organizational readiness, and operational resilience are not downstream consequences but cross-cutting controls that determine whether the transformation is sustainable.

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

6. Limitations of the Review

This review has several limitations. Its scope is intentionally broad, creating heterogeneity that precludes quantitative meta-analysis. The 40-study matrix is a focused corpus designed to represent major clinical, operational, and governance themes rather than every AI application in medicine. PubMed/MEDLINE was the primary structured biomedical index, supplemented by publisher- and index-based searches and citation chaining; therefore, some relevant engineering or health-services studies may not have been captured. The field is also moving rapidly, particularly for generative and agentic AI, so conclusions about emerging applications may change quickly. Many included empirical studies are retrospective, and limitations in the underlying evidence constrain the strength of system-level causal claims. Finally, the concepts of decision intelligence and operational resilience are used as integrative analytical lenses; they are not uniformly defined or measured across the primary literature.

7. Conclusion

AI and health analytics are becoming core components of modern healthcare information and decision infrastructure. Evidence across EHR prediction, sepsis and cardiovascular risk, medical imaging, natural-language processing, federated learning, emergency operations, and forecasting demonstrates substantial technical capability. The strategic challenge is no longer proving that AI can recognize patterns; it is ensuring that those patterns improve decisions, workflows, equity, and outcomes. This review shows that digital transformation occurs when AI is embedded within redesigned decision pathways rather than layered onto existing processes. Decision intelligence links prediction with explanation, accountability, action, and feedback. Operational resilience emerges when these capabilities help healthcare organizations anticipate demand and risk, absorb disruption, adapt resources, recover performance, and learn. Yet AI can also create new dependencies and inequities if deployment is poorly governed.

Accordingly, trustworthy AI requires lifecycle governance, external validation, interoperable data, subgroup evaluation, human-factors design, cybersecurity, monitoring, and workforce readiness. Recent consensus frameworks and 2025-2026 implementation reviews indicate a shift from algorithm-centric research toward deployable socio-technical systems. The next generation of healthcare AI should be judged not only by accuracy, but by whether it produces measurable, safe, equitable, and resilient improvements in care.

References

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine, 1, 39. https://doi.org/10.1038/s41746-018-0040-6

Almufareh, M. F., Tariq, N., Humayun, M., & Almas, B. (2023). A federated learning approach to breast cancer prediction in a collaborative learning framework. Healthcare, 11(24), 3185. https://doi.org/10.3390/healthcare11243185

Al-Nafjan, A., Aljuhani, A., Alshebel, A., Alharbi, A., & Alshehri, A. (2025). Artificial intelligence in predictive healthcare: A systematic review. Journal of Clinical Medicine, 14(19), 6752. https://doi.org/10.3390/jcm14196752

Alanazi, A., Aldakhil, L., Aldhoayan, M., & Aldosari, B. (2023). Machine learning for early prediction of sepsis in intensive care unit (ICU) patients. Medicina, 59(7), 1276. https://doi.org/10.3390/medicina59071276

Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical and healthcare domains: A systematic literature review. Computers in Biology and Medicine, 166, 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., et al. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. BMC Medical Education, 23, 689. https://doi.org/10.1186/s12909-023-04698-z

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

AlShammari, T. (2026). The future of artificial intelligence in emergency medical services by 2030: An international consensus report. Journal of the American College of Emergency Physicians Open, 7(3), 100355. https://doi.org/10.1016/j.acepjo.2026.100355

Amugongo, L. M., Mascheroni, P., Brooks, S., Doering, S., & Seidel, J. (2025). Retrieval augmented generation for large language models in healthcare: A systematic review. PLOS Digital Health, 4(6), e0000877. https://doi.org/10.1371/journal.pdig.0000877

Andersen, A., Huang, R., & Liu, E. J. (2026). The application of artificial intelligence in healthcare practice: A mapping review of systematic reviews. Artificial Intelligence in Medicine, 181, 103495. https://doi.org/10.1016/j.artmed.2026.103495

Armeni, P., Polat, I., De Rossi, L. M., Diaferia, L., Meregalli, S., & Gatti, A. (2022). Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? A critical review. Journal of Personalized Medicine, 12(8), 1255. https://doi.org/10.3390/jpm12081255

Bracken, A., Reilly, C., Feeley, A., Sheehan, E., Merghani, K., & Feeley, I. (2025). Artificial intelligence (AI)-powered documentation systems in healthcare: A systematic review. Journal of Medical Systems, 49(1), 28. https://doi.org/10.1007/s10916-025-02157-4

Celi, L. A., Cellini, J., Charpignon, M. L., et al. (2022). Sources of bias in artificial intelligence that perpetuate healthcare disparities: A global review. PLOS Digital Health, 1(3), e0000022. https://doi.org/10.1371/journal.pdig.0000022

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of explainable AI techniques in healthcare. Sensors, 23(2), 634. https://doi.org/10.3390/s23020634

Chan, H. P., Samala, R. K., Hadjiiski, L. M., & Zhou, C. (2020). Deep learning in medical image analysis. Advances in Experimental Medicine and Biology, 1213, 3-21. https://doi.org/10.1007/978-3-030-33128-3_1

Chang, Y. H., Shih, H. M., Wu, J. E., et al. (2022). Machine learning-based triage to identify low-severity patients with a short discharge length of stay in emergency department. BMC Emergency Medicine, 22, 88. https://doi.org/10.1186/s12873-022-00632-6

Chen, R. J., Wang, J. J., Williamson, D. F. K., et al. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature Biomedical Engineering, 7(6), 719-742. https://doi.org/10.1038/s41551-023-01056-8

Chen, X., Wang, X., Zhang, K., et al. (2022). Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis, 79, 102444. https://doi.org/10.1016/j.media.2022.102444

Cho, S. Y., Kim, S. H., Kang, S. H., et al. (2021). Pre-existing and machine learning-based models for cardiovascular risk prediction. Scientific Reports, 11, 8886. https://doi.org/10.1038/s41598-021-88257-w

Collaco, B. G., Haider, S. A., Prabha, S., et al. (2026). The role of agentic artificial intelligence in healthcare: A scoping review. npj Digital Medicine, 9, 345. https://doi.org/10.1038/s41746-026-02517-5

Coorey, G., Figtree, G. A., Fletcher, D. F., et al. (2022). The health digital twin to tackle cardiovascular disease: A review of an emerging interdisciplinary field. npj Digital Medicine, 5, 126. https://doi.org/10.1038/s41746-022-00640-7

Couture, V., Roy, M. C., Dez, E., Laperle, S., & Bélisle-Pipon, J. C. (2023). Ethical implications of artificial intelligence in population health and the public's role in its governance. Journal of Medical Internet Research, 25, e44357. https://doi.org/10.2196/44357

Cremonesi, F., Planat, V., Kalokyri, V., et al. (2023). The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform. Journal of Biomedical Informatics, 141, 104338. https://doi.org/10.1016/j.jbi.2023.104338

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94-98. https://doi.org/10.7861/futurehosp.6-2-94

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Davenport, T. H., & Glaser, J. P. (2022). Factors governing the adoption of artificial intelligence in healthcare providers. Discover Health Systems, 1, 4. https://doi.org/10.1007/s44250-022-00004-8

Darzidehkalani, E., Ghasemi-Rad, M., & van Ooijen, P. M. A. (2022). Federated learning in medical imaging: Part I: Toward multicentral health care ecosystems. Journal of the American College of Radiology, 19(8), 969-974. https://doi.org/10.1016/j.jacr.2022.03.015

Davis, S., Zhang, J., Lee, I., et al. (2022). Effective hospital readmission prediction models using machine-learned features. BMC Health Services Research, 22, 1415. https://doi.org/10.1186/s12913-022-08748-y

De Hond, A., Raven, W., Schinkelshoek, L., et al. (2021). Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope? International Journal of Medical Informatics, 152, 104496. https://doi.org/10.1016/j.ijmedinf.2021.104496

De Micco, F., Di Palma, G., Ferorelli, D., et al. (2024). Artificial intelligence in healthcare: Transforming patient safety with intelligent systems—A systematic review. Frontiers in Medicine, 11, 1522554. https://doi.org/10.3389/fmed.2024.1522554

Di Martino, F., & Delmastro, F. (2023). Explainable AI for clinical and remote health applications: A survey on tabular and time series data. Artificial Intelligence Review, 56, 5261-5315. https://doi.org/10.1007/s10462-022-10304-3

El-Bouri, R., Taylor, T., Youssef, A., Zhu, T., & Clifton, D. A. (2021). Machine learning in patient flow: A review. Progress in Biomedical Engineering, 3(2), 022002. https://doi.org/10.1088/2516-1091/abddc5

Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29. https://doi.org/10.1038/s41591-018-0316-z

Gao, R., Cheng, W. X., Suganthan, P. N., & Yuen, K. F. (2022). Inpatient discharges forecasting for Singapore hospitals by machine learning. IEEE Journal of Biomedical and Health Informatics, 26(10), 4966-4975. https://doi.org/10.1109/JBHI.2022.3172956

Henzler, D., Schmidt, S., Koçar, A., et al. (2025). Healthcare professionals' perspectives on artificial intelligence in patient care: A systematic review of hindering and facilitating factors on different levels. BMC Health Services Research, 25, 633. https://doi.org/10.1186/s12913-025-12664-2

Holmström, O., Linder, N., Kaingu, H., et al. (2021). Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. JAMA Network Open, 4(3), e211740. https://doi.org/10.1001/jamanetworkopen.2021.1740

Howcroft, A., Bennett-Weston, A., Khan, A., et al. (2025). AI chatbots versus human healthcare professionals: A systematic review and meta-analysis of empathy in patient care. British Medical Bulletin, 156(1), ldaf017. https://doi.org/10.1093/bmb/ldaf017

Huang, W., Ying, T. W., Chin, W. L. C., et al. (2022). Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction. Scientific Reports, 12, 1033. https://doi.org/10.1038/s41598-021-04649-y

Jiang, X., Wang, Y., Pan, Y., & Zhang, W. (2022). Prediction models for sepsis-associated thrombocytopenia risk in intensive care units based on a machine learning algorithm. Frontiers in Medicine, 9, 837382. https://doi.org/10.3389/fmed.2022.837382

Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. Cancers, 15(14), 3608. https://doi.org/10.3390/cancers15143608

Jiang, Z., Bo, L., Xu, Z., et al. (2021). An explainable machine learning algorithm for risk factor analysis of in-hospital mortality in sepsis survivors with ICU readmission. Computer Methods and Programs in Biomedicine, 204, 106040. https://doi.org/10.1016/j.cmpb.2021.106040

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Jung, J., Lee, H., Jung, H., & Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. Heliyon, 9(5), e16110. https://doi.org/10.1016/j.heliyon.2023.e16110

Kather, J. N., Ferber, D., Wiest, I. C., Gilbert, S., & Truhn, D. (2024). Large language models could make natural language again the universal interface of healthcare. Nature Medicine, 30(10), 2708-2710. https://doi.org/10.1038/s41591-024-03199-w

Katirai, A. (2023). The ethics of advancing artificial intelligence in healthcare: Analyzing ethical considerations for Japan's innovative AI hospital system. Frontiers in Public Health, 11, 1142062. https://doi.org/10.3389/fpubh.2023.1142062

Khullar, D., Wang, X., & Wang, F. (2024). Large language models in health care: Charting a path toward accurate, explainable, and secure AI. Journal of General Internal Medicine, 39(7), 1239-1241. https://doi.org/10.1007/s11606-024-08657-2

Kim, M., Sohn, H., Choi, S., & Kim, S. (2023). Requirements for trustworthy artificial intelligence and its application in healthcare. Healthcare Informatics Research, 29(4), 315-322. https://doi.org/10.4258/hir.2023.29.4.315

King, Z., Farrington, J., Utley, M., et al. (2022). Machine learning for real-time aggregated prediction of hospital admission for emergency patients. npj Digital Medicine, 5, 104. https://doi.org/10.1038/s41746-022-00649-y

Kiser, A. C., Eilbeck, K., Ferraro, J. P., et al. (2022). Standard vocabularies to improve machine learning model transferability with electronic health record data: Retrospective cohort study using health care-associated infection. JMIR Medical Informatics, 10(8), e39057. https://doi.org/10.2196/39057

Kondylakis, H., Osuala, R., Puig-Bosch, X., et al. (2026). A review of methods for trustworthy AI in medical imaging: The FUTURE-AI guidelines. IEEE Journal of Biomedical and Health Informatics, 30(3), 2299-2315. https://doi.org/10.1109/JBHI.2025.3614546

Koo, T. H., Zakaria, A. D., Ng, J. K., & Leong, X. B. (2024). Systematic review of the application of artificial intelligence in healthcare and nursing care. Malaysian Journal of Medical Sciences, 31(5), 135-142. https://doi.org/10.21315/mjms2024.31.5.9

Lanfranchi, G., Gembillo, G., Santoro, D., & Villari, M. (2026). Artificial intelligence adoption in healthcare: A systematic review of implementation challenges and health services implications. BMC Health Services Research. https://doi.org/10.1186/s12913-026-14960-x

Lee, H., Chai, Y. J., Joo, H., et al. (2021). Federated learning for thyroid ultrasound image analysis to protect personal information: Validation study in a real health care environment. JMIR Medical Informatics, 9(5), e25869. https://doi.org/10.2196/25869

Lekadir, K., Frangi, A. F., Porras, A. R., et al. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ, 388, e081554. https://doi.org/10.1136/bmj-2024-081554

Liao, F., Adelaine, S., Afshar, M., & Patterson, B. W. (2022). Governance of clinical AI applications to facilitate safe and equitable deployment in a large health system: Key elements and early successes. Frontiers in Digital Health, 4, 931439. https://doi.org/10.3389/fdgth.2022.931439

Liu, W., et al. (2023). Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: A systematic review and meta-analysis. European Heart Journal - Quality of Care and Clinical Outcomes, 9(4), 310-322. https://doi.org/10.1093/ehjqcco/qcad017

Lo, Y. T., Liao, J. C., Chen, M. H., Chang, C. M., & Li, C. T. (2021). Predictive modeling for 14-day unplanned hospital readmission risk by using machine learning algorithms. BMC Medical Informatics and Decision Making, 21, 288. https://doi.org/10.1186/s12911-021-01639-y

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). Computer Methods and Programs in Biomedicine, 226, 107161. https://doi.org/10.1016/j.cmpb.2022.107161

Lungren, M. P., Fishman, E. K., Chu, L. C., Rizk, R. C., & Rowe, S. P. (2024). More is different: Large language models in health care. Journal of the American College of Radiology, 21(7), 1151-1154. https://doi.org/10.1016/j.jacr.2023.11.021

Machado, T. M., & Berssaneti, F. T. (2023). Literature review of digital twin in healthcare. Heliyon, 9(9), e19390. https://doi.org/10.1016/j.heliyon.2023.e19390

McKay, F., Williams, B. J., Prestwich, G., Treanor, D., & Hallowell, N. (2022). Public governance of medical artificial intelligence research in the UK: An integrated multi-scale model. Research Involvement and Engagement, 8, 21. https://doi.org/10.1186/s40900-022-00357-7

Moor, M., Rieck, B., Horn, M., Jutzeler, C. R., & Borgwardt, K. (2021). Early prediction of sepsis in the ICU using machine learning: A systematic review. Frontiers in Medicine, 8, 607952. https://doi.org/10.3389/fmed.2021.607952

Morley, J., Murphy, L., Mishra, A., Joshi, I., & Karpathakis, K. (2022). Governing data and artificial intelligence for health care: Developing an international understanding. JMIR Formative Research, 6(1), e31623. https://doi.org/10.2196/31623

Narmadha, K., & Varalakshmi, P. (2022). Federated learning in healthcare: A privacy preserving approach. Studies in Health Technology and Informatics, 294, 194-198. https://doi.org/10.3233/SHTI220436

Nerella, S., Bandyopadhyay, S., Zhang, J., et al. (2024). Transformers and large language models in healthcare: A review. Artificial Intelligence in Medicine, 154, 102900. https://doi.org/10.1016/j.artmed.2024.102900

Nguyen, M., Corbin, C. K., Eulalio, T., et al. (2021). Developing machine learning models to personalize care levels among emergency room patients for hospital admission. Journal of the American Medical Informatics Association, 28(11), 2423-2432. https://doi.org/10.1093/jamia/ocab118

Njei, B., Al-Ajlouni, Y. A., Kanmounye, U. S., et al. (2026). Artificial intelligence agents in healthcare research: A scoping review. PLOS ONE, 21(2), e0342182. https://doi.org/10.1371/journal.pone.0342182

O'Connor, S., & Booth, R. G. (2022). Algorithmic bias in health care: Opportunities for nurses to improve equality in the age of artificial intelligence. Nursing Outlook, 70(6), 780-782. https://doi.org/10.1016/j.outlook.2022.09.003

Ordikhani, M., Saniee Abadeh, M., Prugger, C., et al. (2022). An evolutionary machine learning algorithm for cardiovascular disease risk prediction. PLOS ONE, 17(7), e0271723. https://doi.org/10.1371/journal.pone.0271723

Patel, D., Cheetirala, S. N., Raut, G., et al. (2022). Predicting adult hospital admission from emergency department using machine learning: An inclusive gradient boosting model. Journal of Clinical Medicine, 11(23), 6888. https://doi.org/10.3390/jcm11236888

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1, 18. https://doi.org/10.1038/s41746-018-0029-1

Rogers, W. A., Draper, H., & Carter, S. M. (2021). Evaluation of artificial intelligence clinical applications: Detailed case analyses show value of healthcare ethics approach in identifying patient care issues. Bioethics, 35(7), 623-633. https://doi.org/10.1111/bioe.12885

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Sahoo, R. K., Sahoo, K. C., Negi, S., Baliarsingh, S. K., Panda, B., & Pati, S. (2025). Health professionals' perspectives on the use of artificial intelligence in healthcare: A systematic review. Patient Education and Counseling, 134, 108680. https://doi.org/10.1016/j.pec.2025.108680

Schmidt, J., Carter, A. W., McGuire, A., Mossialos, E., & van Kessel, R. (2026). A systematic review of economic evidence of artificial intelligence in healthcare. Health Policy, 172, 105715. https://doi.org/10.1016/j.healthpol.2026.105715

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: A structured literature review. BMC Medical Informatics and Decision Making, 21, 125. https://doi.org/10.1186/s12911-021-01488-9

Seibert, K., Domhoff, D., Fürstenau, D., et al. (2023). Exploring needs and challenges for AI in nursing care: Results of an explorative sequential mixed methods study. BMC Digital Health, 1, 13. https://doi.org/10.1186/s44247-023-00015-2

Shi, H., Huang, J., Yang, J., & Han, M. (2026). Barriers and facilitators to patient acceptance of artificial intelligence in health care: Systematic review. Journal of Medical Internet Research, 28, e80581. https://doi.org/10.2196/80581

Spear, J., Ehrenfeld, J. M., & Miller, B. J. (2023). Applications of artificial intelligence in health care delivery. Journal of Medical Systems, 47, 121. https://doi.org/10.1007/s10916-023-02018-y

Starke, G., Gille, F., Termine, A., et al. (2025). Finding consensus on trust in AI in health care: Recommendations from a panel of international experts. Journal of Medical Internet Research, 27, e56306. https://doi.org/10.2196/56306

Steiner, D. F., Nagpal, K., Sayres, R., et al. (2020). Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. JAMA Network Open, 3(11), e2023267. https://doi.org/10.1001/jamanetworkopen.2020.23267

Sun, T., He, X., & Li, Z. (2023). Digital twin in healthcare: Recent updates and challenges. Digital Health, 9, 20552076221149651. https://doi.org/10.1177/20552076221149651

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Torres-Velázquez, M., Chen, W. J., Li, X., & McMillan, A. B. (2021). Application and construction of deep learning networks in medical imaging. IEEE Transactions on Radiation and Plasma Medical Sciences, 5(2), 137-159. https://doi.org/10.1109/TRPMS.2020.3030611

Tschoellitsch, T., Seidl, P., Böck, C., et al. (2023). Using emergency department triage for machine learning-based admission and mortality prediction. European Journal of Emergency Medicine, 30(6), 408-416. https://doi.org/10.1097/MEJ.0000000000001068

Tveit, J., Aurlien, H., Plis, S., et al. (2023). Automated interpretation of clinical electroencephalograms using artificial intelligence. JAMA Neurology. https://doi.org/10.1001/jamaneurol.2023.1645

Vântu, A., Vasilescu, A., & Băicoianu, A. (2023). Medical emergency department triage data processing using a machine-learning solution. Heliyon, 9(8), e18402. https://doi.org/10.1016/j.heliyon.2023.e18402

Vollmer, S., Mateen, B. A., Bohner, G., et al. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ, 368, l6927. https://doi.org/10.1136/bmj.l6927

Waldock, W. J., Zhang, J., Guni, A., Nabeel, A., Darzi, A., & Ashrafian, H. (2024). The accuracy and capability of artificial intelligence solutions in health care examinations and certificates: Systematic review and meta-analysis. Journal of Medical Internet Research, 26, e56532. https://doi.org/10.2196/56532

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Wang, D., Li, J., Sun, Y., et al. (2021). A machine learning model for accurate prediction of sepsis in ICU patients. Frontiers in Public Health, 9, 754348. https://doi.org/10.3389/fpubh.2021.754348

Wang, L., Wan, Z., Ni, C., et al. (2024). Applications and concerns of ChatGPT and other conversational large language models in health care: Systematic review. Journal of Medical Internet Research, 26, e22769. https://doi.org/10.2196/22769

Xi, Y., Wang, H., & Sun, N. (2022). Machine learning outperforms traditional logistic regression and offers new possibilities for cardiovascular risk prediction: A study involving 143,043 Chinese patients with hypertension. Frontiers in Cardiovascular Medicine, 9, 1025705. https://doi.org/10.3389/fcvm.2022.1025705

Yang, X., Chen, A., PourNejatian, N., et al. (2022). A large language model for electronic health records. npj Digital Medicine, 5, 194. https://doi.org/10.1038/s41746-022-00742-2

Younis, H. A., Eisa, T. A. E., Nasser, M., et al. (2024). A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: Applications, considerations, limitations, motivation and challenges. Diagnostics, 14(1), 109. https://doi.org/10.3390/diagnostics14010109

Yu, J. Y., Xie, F., Liu, N., et al. (2022). An external validation study of the Score for Emergency Risk Prediction (SERP), an interpretable machine learning-based triage score for the emergency department. Scientific Reports, 12, 17466. https://doi.org/10.1038/s41598-022-22233-w

Yu, M., Tang, A., Brown, K., et al. (2021). Integrating artificial intelligence in bedside care for COVID-19 and future pandemics. BMJ, 375, e068197. https://doi.org/10.1136/bmj-2021-068197

Zhang, J., & Zhang, Z. M. (2023). Ethics and governance of trustworthy medical artificial intelligence. BMC Medical Informatics and Decision Making, 23, 7. https://doi.org/10.1186/s12911-023-02103-9

Zhang, Y., Hu, J., Hua, T., et al. (2023). Development of a machine learning-based prediction model for sepsis-associated delirium in the intensive care unit. Scientific Reports, 13, 12697. https://doi.org/10.1038/s41598-023-38650-4