

Intelligent Healthcare: AI and Machine Learning in Diagnosis, Precision Medicine, and the Pursuit of Explainable, Trustworthy Clinical Systems

Ekta R. Kaneriya¹

ekta.kaneriya14@gmail.com

Atmiya University

Maulik K. Shah²

Bhakta Kavi Narsinh Mehta University

Krishna H. Sapariya

Bhakta Kavi Narsinh Mehta University

Research Article

Keywords: Artificial Intelligence, Machine Learning, Healthcare, Deep Learning, Medical Diagnosis, Precision Medicine, Medical Imaging, Predictive Analytics, Electronic Health Records, Explainable AI

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11090888/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence (AI) and machine learning (ML) are increasingly used across healthcare to analyze complex clinical data, support diagnosis, guide precision medicine, accelerate drug discovery, and enable remote monitoring. This review examines established AI/ML techniques and their current clinical and commercial applications while considering the requirements for trustworthy adoption. Evidence reviewed across medical imaging, electronic health records, drug discovery, precision medicine, and remote patient monitoring shows broad use of supervised, unsupervised, deep-learning, hybrid, and ensemble approaches. Reported benefits include improved diagnostic and predictive performance, earlier risk identification, and support for individualized treatment. Recurring barriers include data privacy, model opacity, computational demands, algorithmic bias, limited external validation, workflow integration, and workforce readiness. Explainable AI, privacy-preserving learning, human-in-the-loop oversight, and evolving regulatory frameworks are prominent responses to these challenges. AI/ML is moving from experimental development toward practical healthcare use, but dependable adoption requires more than predictive accuracy. Robust external validation, transparent and interpretable models, privacy protection, equitable datasets, appropriate regulation, and effective clinician-AI collaboration are central to building trustworthy clinical systems.

1. Background

Diagnosis and treatment remain fundamentally grounded in clinical expertise, but modern healthcare systems also generate large volumes of imaging, laboratory, genomic, narrative, and sensor-derived data. AI and ML can augment clinicians by organizing and analyzing these data at a scale that is difficult to achieve through manual review alone (Dixon et al., 2024; Alhumaidi et al., 2025).

The transition has accelerated over the past decade. Convolutional neural networks and related deep-learning models are widely studied for medical image classification and segmentation, natural-language-processing systems can extract structured information from clinical text, and generative models are being explored for drug discovery and molecular design (Sistaninejhad et al., 2023; Berghout, 2025; Ferreira & Carneiro, 2025; Niazi, 2025).

However, the growing impact of AI/ML in healthcare has also intensified scrutiny of these systems. A model that performs well on a benchmark dataset may not remain reliable in clinical practice, where patient populations, imaging equipment, documentation patterns, and care processes can differ across institutions. Concerns about data privacy, algorithmic bias, legal responsibility, and the opacity of complex models therefore remain central to the discussion (Liu et al., 2025; Rosenbacke et al., 2024; Mohammed & Malhotra, 2025).

Throughout the article, the notion of “case studies” is used in a relatively literal sense, with the algorithms described not just abstractly but also in the context of an actual deployment or evaluated system, as reported in peer-reviewed reviews and meta-analyses, primarily from the past 3–5 years. It's a conscious decision. When it comes to AI in healthcare, there's a vast amount of proof-of-concept work

that exists in the lab, and separating out that from systems that have been validated against real patient outcomes, cleared by a regulator, or that have been adopted into the workflow of a hospital is one of the more valuable services that a review article like this one can offer.

This article is arranged in that fashion. It starts by presenting the fundamental types of algorithms: supervised learning, unsupervised learning and deep learning, which are used throughout the rest of the article. It then explores five areas where AI/ML has had a tangible clinical impact—medical imaging, electronic health record analysis, drug discovery, precision medicine, and remote patient monitoring—and highlights examples of its use (where possible) instead of hypothetical scenarios. The second half of the article focuses on the challenges of getting promising research to reliable clinical infrastructure: privacy exposure, model opacity, computational cost and ethical hazard, followed by an overview of the regulatory framework and explainable AI techniques that are slowly being developed to address these challenges. The article ends with a discussion of the future directions of the field.

2. Methods

This article uses a narrative review approach to synthesize evidence on artificial intelligence (AI) and machine learning (ML) in healthcare. Because the original literature collection was not prospectively registered or conducted as a formal systematic review, the search process was retrospectively documented to improve search transparency and reproducibility. The literature supporting the review was verified using PubMed, Google Scholar, and Semantic Scholar. The core search terms were combined as follows: ("artificial intelligence" OR "machine learning") AND (healthcare OR "health care" OR clinical) AND ("medical imaging" OR "electronic health records" OR EHR OR "drug discovery" OR "precision medicine" OR "remote patient monitoring" OR diagnosis). Domain-specific searches using combinations of artificial intelligence, machine learning, and each of the five application areas were also used where needed to locate relevant literature. The primary focused literature coverage was 2021–2026, while earlier studies were retained when they provided important methodological, historical, or conceptual context. The last literature verification was performed on 5 October 2026. English-language publications were prioritized. Eligible evidence included peer-reviewed reviews, systematic reviews, meta-analyses, and relevant clinical or commercial case examples addressing healthcare AI/ML, as well as selected earlier or non-journal sources when they provided directly relevant methodological, technical, or regulatory context. Duplicate and clearly irrelevant records were excluded. Titles and abstracts were screened for relevance, followed by full-text assessment of potentially relevant publications. The final evidence set was organized across five application domains: medical imaging, electronic health record analytics, drug discovery, precision medicine, and remote patient monitoring. The synthesis also examines privacy, interpretability, computational requirements, ethics and bias, generalizability, workflow integration, regulation, and explainable AI. Because the article is a literature-based review and does not involve primary human or animal data collection, no participant recruitment, intervention, or statistical power calculation was performed.

3. Results: Foundational AI/ML Techniques in Healthcare

In order to explore specific clinical applications, it is useful to develop the language of the methods that underlie the system, as the selection of methods influences the capabilities and limitations of a system in a medical context.

3.1 Supervised Learning

Supervised learning is the most common type of learning used in clinical AI. The model is trained with examples, which are already labeled with the “right answer”, such as a mammogram labeled as malignant or benign or a lab panel linked to its eventual diagnosis, and learns to map new, unlabeled inputs to the same categories or continuous outcomes. Structured tabular data like EHR variables is still widely used in this area and is where logistic regression, random forests, gradient-boosted trees (XGBoost, LightGBM) and support vector machines are still widely used, as they tend to be faster to train, audit and less data hungry than deep networks. Adverse-drug-event prediction from EHR data has been a popular application for reviewing the performance of different model families; random forest and gradient-boosting methods are consistently among the most frequently reported and best-performing model families (Hu et al., 2024; Deimazar & Sheikhtaheri, 2023).

3.2 Unsupervised Learning

When outcomes are unlabeled or the objective is to discover structure in the data, unsupervised methods can be useful. Clustering algorithms such as k-means, hierarchical clustering, and DBSCAN can group patients according to similarities across multiple variables, while dimensionality-reduction methods such as principal component analysis (PCA) and t-SNE can support visualization and exploration of high-dimensional genomic or proteomic data. Related anomaly-detection approaches are also used in remote monitoring to identify physiological readings that deviate from an individual baseline (Alhumaidi et al., 2025; Shaik et al., 2023).

3.3 Deep Learning

Over the last decade, most of the medical AI breakthroughs have been achieved with deep learning architectures, which are artificial neural networks with numerous layers. CNNs are the most effective for image-based tasks as the layers of filters are well suited to capturing spatial features like shape of a nodule or texture of a lesion; architectures like U-Net are commonly used for segmentation, and transfer learning from large pretrained networks can help overcome the comparatively limited size of most medical imaging datasets (Sistaninejhad et al., 2023; Berghout, 2025). Architectures with a recurrent structure and, more recently, transformer models are used to process sequential data, like time-series of vital data or free-text clinical notes, where the sequence of data is meaningful. Drug-candidate generation and synthetic data generation have increasingly been brought into the fold of generative approaches, such as variational autoencoders, generative adversarial networks, and diffusion models, and graph neural networks are starting to become more popular for directly modeling molecular structures and protein interactions (Niazi, 2025; Khare et al., 2026).

3.4 Hybrid and Ensemble Approaches

In practice, clinical AI systems often combine multiple model components rather than relying on a single algorithm. A deep-learning feature extractor may feed a simpler classifier, or several model families may be combined in an ensemble to improve predictive stability. For structured EHR data, tree-based ensembles and boosting methods can provide competitive performance together with feature-importance summaries that are easier to inspect than the internal parameters of a large neural network (Liu et al., 2025; Hu et al., 2024).

3.5 Evaluation and Validation Considerations

Model performance must be interpreted using multiple metrics. Area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and positive predictive value (PPV) capture different aspects of discrimination and clinical usefulness. In low-prevalence settings, high AUC and specificity can coexist with relatively low PPV because the target event is rare. External validation is therefore essential because models that perform well on development data may underperform at another hospital with different patient populations, equipment, or documentation practices (Hu et al., 2024; Liu et al., 2025; Deimazar & Sheikhtaheri, 2023).

These families and evaluation practices have their respective advantages and disadvantages in terms of predictive power, robustness, and interpretability, which will continually appear throughout the rest of this article. The three core AI/ML technique families and representative algorithms used in healthcare are summarized in Fig. 1.

3.6 Applications of AI/ML Across Healthcare Domains

Figure 2 provides an overview of five application domains explored in this article, each linked to the common set of AI/ML techniques described above.

3.6.1 Medical Imaging and Diagnostic Radiology

Medical imaging is perhaps the field where deep learning has come the furthest, in part because imaging data is naturally well suited to convolutional architectures, and in part because there are pattern recognition tasks in medical imaging, such as classification of images in radiology, pathology, and dermatology, which are well suited to a mapping of image to class. The field has been reviewed systematically, with consistent improvements made by CNN-based systems in organ segmentation, lesion detection, and disease classification; and the use of pretrained backbones and generative adversarial networks is steadily becoming more popular to squeeze better performance from comparatively small labeled datasets (Sistaninejhad et al., 2023).

One such example that has been studied is breast cancer screening. For mammography and ultrasound, critical reviews of computer-aided diagnosis systems have shown that deep learning classifiers have

increasingly eliminated the need for hand-engineered feature extraction, and directly learned discriminative patterns from the pixel data, leading to screening tools that support radiologists to identify suspicious lesions for further review (Jiménez-Gaona et al., 2020). The same goes for brain tumor detection: in a study of over one hundred papers published from 2019 to 2024, CNNs, and their transfer-learning cousins are the most prevalent approaches, while generative adversarial networks and autoencoders are used to supplement limited training data, and recurrent networks are used when temporal imaging sequences are available (Berghout, 2025).

The COVID-19 pandemic has created an extraordinary case study in fast deployment of imaging-AI. Studies of classification pipelines (both chest X-ray and CT) demonstrate the ability to reach an accuracy of more than 99 percent on curated datasets when distinguishing between COVID-19 pneumonia and other pneumonia and normal chest X-ray or CT (Bharati et al., 2021). The more recent research on brain-image synthesis has taken a step further towards generative methods, employing deep networks to reconstruct or enhance brain images, and suggesting transformer-based architectures as the next promising step in the evolution of brain-image synthesis techniques (Singh et al., 2023).

In addition to these "on the front lines" examples, imaging AI has also started to venture beyond traditional radiology. When a new imaging modality is introduced to the clinic, the same deep-learning toolkit is typically applied to it rather than a new modeling paradigm being developed from the ground up (Sistaninejad et al., 2023).

One common theme in all these imaging case studies is the discrepancy between the performance of the algorithms in the benchmarks and the actual clinical tools that are available for deployment. While high classification accuracy on a curated, retrospective set is a prerequisite, it is not sufficient on its own. The systems that have actually reached clinical use typically add safeguards that keep a human reviewer in the loop, even when the underlying algorithm is highly accurate, for example by requiring a radiologist's sign-off before a result is entered into the patient chart (Mohammed & Malhotra, 2025).

3.6.2 Electronic Health Record Analysis and Predictive Analytics

Imaging AI is about pixels, while EHR-based AI is about the vast, disorganized patient health history of diagnoses, medication lists, lab trends, vital signs and free-text notes. The promise is simple: By building predictive models on top of EHRs, this information can be gathered without asking patients to provide anything extra, and it can give early warnings without any extra effort.

One of the most well validated use cases is cardiovascular risk prediction. In a 2025 systematic review and meta-analysis of the results of 20 studies, machine learning models (especially random forest and deep-learning models) were shown to be more discriminatory than conventional risk scores like QRISK3 and ASCVD, with the pooled discrimination statistics being significantly higher than the conventional scores they were compared to (Liu et al., 2025). Another often mentioned example is sepsis detection: Systems that continuously analyze data from the EHR and trigger alerts when sepsis is suspected have

been embedded directly into hospital workflows, rather than relying on routine rounds to detect it (Dixon et al., 2024).

EHR-based models are also being used defensively, to detect issues early on and before they get worse. Random forest and other ensemble methods are the subject of dozens of studies reported in systematic reviews for predicting adverse drug events, with good discrimination but low positive predictive value, for the rare events for which they are frequently used to predict (Hu et al., 2024; Deimazar & Sheikhtaheri, 2023). In a parallel line of research, other researchers have been using similar techniques to detect and screen EHRs for the early detection of falls, pressure injuries, and other preventable harms (Deimazar & Sheikhtaheri, 2023). In psychiatry, models trained on structured and unstructured fields of EHR have been used to flag adolescents at moderate to good discrimination who are at short-term risk of a mental health crisis, but with the clear recognition that there is much more to be done before this kind of application can be widely adopted (Al-Juhani et al., 2025). In a general mental-health service population, a gradient-boosting model predicted crisis onset within 28 days with an AUROC of 0.797 (Garriga et al., 2022).

Figure 3. Reported discrimination values with source-specific uncertainty intervals. CVD risk values are pooled AUCs with 95% CIs (Liu et al., 2025); the mental-health crisis value is a single-study AUROC (patients aged 16 years and older in a general mental-health service) with its reported 95% CI (Garriga et al., 2022); and the adverse-drug-event value is the mean AUC $\pm$ SD across studies (Hu et al., 2024). The values are shown descriptively and should not be interpreted as a common pooled comparison.

In addition to these individually studied use cases, a larger literature systematically reviews machine learning using real-world data for disease prediction and management. That literature highlights substantial variation in study design, outcome definitions, reporting standards, and external validation, which limits direct comparison across studies (Alhumaidi et al., 2025).

3.6.3 AI-Driven Drug Discovery

The traditional drug development process is very time consuming and costly, with a compound taking 10 years or more from identification of a target to market approval and the vast majority of compounds failing at various stages of the process (Niazi, 2025). AI/ML is now making its way into almost every part of that pipeline to try and shorten the time and boost the likelihood of success.

A large-scale review covering more than three hundred articles published between 2000 and 2022 maps how deep learning has been applied to drug-target interaction prediction, drug-drug similarity modeling, and side-effect forecasting, noting that recurrent networks, autoencoders, and message-passing graph networks have each found a niche depending on whether the input is a molecular sequence, a 2D structure, or a graph representation of a compound (Askr et al., 2023). More recent reviews extend this picture into generative chemistry, describing how diffusion models and other generative architectures are now used to design entirely novel candidate molecules with specified properties rather than merely screening existing compound libraries, alongside early experimentation with autonomous “agentic”

systems that can plan and execute portions of the discovery workflow with reduced human intervention (Niazi, 2025).

A systematic review focused specifically on studies that carried their computational predictions through to in vivo animal validation found that only a relatively small subset of published deep-learning drug-discovery work reaches that stage, which is a useful reminder that computational promise and biological confirmation are not the same thing: this gap between simulated and validated performance is one of the field's recurring caveats (Koutroumpa et al., 2023). Recent reviews also describe the expanding role of AI across the drug-discovery pipeline (Ferreira & Carneiro, 2025). Interpretability is also a specific line of research within the field of drug discovery (Jiménez-Luna et al., 2020): Regulators and chemists alike are interested in understanding why a generative model suggested a particular molecular structure, so explainable AI methods from other fields are now increasingly being integrated directly into the drug discovery process, rather than added on as an afterthought.

The driving force behind almost all of this work is cost and/or time constraints. When considering the potential for AI-driven approaches, it's important to set these expectations in the context of the development costs, which have been estimated to be more than two and a half billion dollars per approved drug, and the time, which has been estimated as ten to fifteen years, for each compound in the pipeline, small savings in hit-identification accuracy or early toxicity prediction can have a significant impact downstream (Niazi, 2025). In this economic context, it's no wonder that pharmaceutical companies have been relatively quick to embrace AI/ML compared to other areas of healthcare: The benefits in an economic sense such as shorter development timelines or fewer late-stage failures, are more easily measured in dollars and development-months, and are more likely to be tangible.

3.6.4 Precision Medicine

Precision medicine aims to move away from population-averaged treatment protocols toward therapies calibrated to an individual patient's genetic, molecular, and clinical profile, and AI/ML is the computational engine that makes integrating those disparate data types feasible at scale. Multi-modal models now combine genomic sequencing data, imaging findings, and EHR variables within a single predictive framework, an approach that has, for example, allowed deep-learning facial-analysis tools to assist in recognizing syndromic genetic disorders from craniofacial photographs and enabled natural language processing systems to mine unstructured clinical notes for rare-disease indicators that would be easy for a human reviewer to overlook amid routine documentation (Khan et al., 2026).

Oncology has been an especially active testing ground. Reviews of AI-assisted precision oncology describe platforms that match patients to therapies based on multi-omic profiles spanning proteomic and genomic data, with academic groups developing classification tools intended to route autoimmune and cancer patients toward the treatment most likely to work for their specific molecular subtype rather than a generic first-line protocol (Khare et al., 2026). Broader narrative reviews of the field emphasize that realizing this potential at scale will depend on generating more robust clinical evidence, adopting privacy-preserving techniques such as federated learning so that models can be trained across

institutions without centralizing sensitive genomic data, and building governance structures sturdy enough to keep pace with how quickly the underlying models are evolving (Zeng et al., 2026).

A recurring caution across the precision-medicine literature is that molecular sophistication does not automatically translate into equitable access. Genomic reference databases have historically overrepresented patients of European ancestry, and a model trained predominantly on that data can produce less reliable risk estimates or treatment recommendations for patients from underrepresented ancestral backgrounds. Addressing this is less a modeling problem than a data-collection and inclusion problem, and it is one of the reasons several recent reviews list cohort diversification as a precondition for precision medicine living up to its name rather than becoming, in practice, precision medicine for only a subset of the population it claims to serve (Zeng et al., 2026).

3.6.5 Remote Patient Monitoring

The last use case that we'll be exploring is remote patient monitoring (RPM), which takes AI/ML outside the hospital. Wearable sensors and home devices are now constantly collecting data like heart rate, oxygen saturation, respiratory rate and activity level, and AI models are analyzing this data stream, and differentiating between a normal variation and a real warning sign. A thorough literature survey of AI-driven RPM architectures has been conducted, which has advanced from basic threshold alarms to more sophisticated ones, such as activity classification models that learn an individual's normal movement patterns, federated learning architectures that enable personalised monitoring without sharing raw data across patients, and reinforcement learning-based systems that are adaptable over time to the individual's changing baseline (Shaik et al., 2023).

Some of the more developed deployments are in the area of cardiac care. Reviews of wearable-enabled heart-failure monitoring focus on the use of AI/ML algorithms on time-series physiological data to detect clinical deterioration in a timely fashion to prevent hospital admission (Gautam et al., 2022), and prospective studies, such as the SMART-CARE cohort study, are currently underway to determine whether continuous, AI-processed wearable data can meaningfully reduce hospitalizations relative to standard care (Ciccarelli et al., 2025). From the regulatory point of view, an analysis of FDA-cleared RPM devices revealed that dozens of AI-powered monitoring devices have already been cleared in the U.S. – most through the agency's 510(k) pathway – suggesting that this is one of the more advanced areas of clinical AI from a regulatory standpoint, even as the regulatory framework for adaptive algorithms is still being developed (Dubey & Tiwari, 2023).

3.6.6 Cross-Domain Observations

Looking across the five domains surveyed above, several patterns recur. First, imaging and drug discovery rely heavily on deep architectures because their inputs (pixels and molecular graphs) are well suited to representation learning, whereas EHR and remote-monitoring applications more often use ensemble or hybrid pipelines for structured, tabular, or time-series data. Second, a large body of retrospective single-site research reports strong performance, while fewer studies provide prospective or multi-site validation or document routine clinical use. Third, data scarcity and class imbalance remain

persistent constraints, encouraging the use of transfer learning, synthetic augmentation, and related strategies (Liu et al., 2025; Niazi, 2025; Shaik et al., 2023).

4. Discussion

Figure 4 groups the four challenge categories discussed below into a single overview before each is examined in turn.

The clinical successes described above should not obscure the unresolved problems that remain between promising research and dependable, everyday medical infrastructure.

4.1 Data Privacy

Every application discussed in this article depends on access to sensitive patient data (imaging studies, genomic sequences, medication histories, and continuous physiological streams), and that dependency creates real exposure. Even de-identified datasets can sometimes be re-linked to individuals when combined with other available information, and the multi-institutional data sharing that many AI models need in order to generalize well runs directly against the instinct of health systems to keep patient data contained within their own walls. There are some techniques like federated learning, differential privacy and synthetic data generation that have been proposed as partial solutions and are currently being developed but not formalized (Khalid et al., 2023; Naik et al., 2024).

4.2 The Black-Box Problem

In fact, many of the models that achieve the best performance on standard tasks are so structurally complex that they are hard for a human to understand. A decision tree can express its decisions as simple, human-readable rules, whereas a convolutional network with millions of parameters cannot. That opacity is not something that can be tolerated in a field where a wrong decision could lead to the wrong diagnosis of cancer or a wrong recommendation for treatment. This “black-box” nature is one of the major challenges for the clinical adoption of medical image analysis systems, as it may lead to serious medicolegal consequences, and clinicians are certainly hesitant to entrust a system whose reasoning they cannot check (Muhammad & Bendeche, 2024).

4.3 Computational Resources

Modern deep-learning systems can require substantial computing infrastructure, specialized hardware, large-scale storage, and technical personnel for training and maintenance, particularly for large imaging and genomic models. These resource requirements may disadvantage smaller hospitals and health systems with limited infrastructure and can complicate local deployment of models developed elsewhere (Niazi, 2025; Mohammed & Malhotra, 2025).

4.4 Ethical Concerns and Algorithmic Bias

A model, built on past medical information, will have the same biases as the medical information. With an under-represented population in the training set, the model is likely to be less effective at predicting the outcomes of a patient from that demographic group, which may be those who are already most likely to experience poor care. Along with the issues of legal liability when an AI-assisted decision leads to a patient harm, which could be plausibly laid at the developer of the system, the treating clinician, or the institution that deployed the system, and the lack of clarity in the law in most jurisdictions, algorithmic bias, due to non-representative training data, is one of the most common ethical concerns cited in the literature in reviews of implementation barriers in machine-learning-based healthcare systems (Mohammed & Malhotra, 2025; Bender et al., 2025). A related challenge is informed consent, as a lot of patient information is used for the creation of AI models in ways that consent forms were not intended to foresee (Bender et al., 2025).

4.5 Generalizability and Workflow Integration

A model that works well under controlled research conditions does not necessarily become a tool that clinicians will use. Once a model is deployed in a real production environment, there are many subtle factors that can creep into the data and compromise the accuracy of the model, such as the differences in case-mix, differences in the calibration of imaging equipment, differences in laboratory reference ranges, and differences in how a hospital's staff enter the data into the EHR. The lack of external validation is one of the most frequently cited limitations in the published literature, either for cardiovascular risk prediction or real-world-data applications, with many models having impressive statistics, but never tested outside the environment in which they were developed (Liu et al., 2025; Alhumaidi et al., 2025). Workflow integration makes it worse: If alerts occur too frequently, clinicians will become immune to them (often referred to as "alert fatigue") and if they occur too infrequently, they may not be noticed when time to act is right. The design of the sensitivity, timing and presentation of an AI-generated recommendation is just as much a human-factors design challenge as it is a modelling challenge and is often overlooked in the academic literature when compared with the predictive power of the model (Mohammed & Malhotra, 2025).

4.6 Workforce Readiness

Workforce readiness is another important barrier. Clinicians may need training to interpret AI-generated risk scores, recognize model uncertainty, decide when to override an automated recommendation, and communicate AI-supported assessments appropriately to patients. Implementation reviews therefore emphasize education, governance, and human oversight as complements to technical model development (Mohammed & Malhotra, 2025).

4.7 Regulatory Landscape and Explainable AI

The challenges listed above are not only technical in nature, so is the answer to them, which is becoming increasingly regulatory and methodological.

Regulatory oversight of AI-enabled healthcare technologies is evolving alongside the technology. Reviews of AI-enabled remote-monitoring devices in the United States report substantial use of the FDA 510(k) pathway, which is based on substantial equivalence to a predicate device (Dubey & Tiwari, 2023). Broader implementation research also emphasizes the need for clinical validation, lifecycle monitoring, transparency, accountability, and governance as AI systems evolve (Mohammed & Malhotra, 2025).

The black-box problem described above has motivated the development and application of Explainable AI (XAI). Systematic reviews of healthcare XAI describe model-agnostic methods such as SHAP and LIME for structured data and image-based techniques such as Grad-CAM for visual explanations. However, explanation methods do not automatically create appropriate clinical trust: empirical evidence shows that explanations can increase or decrease clinician confidence depending on how they are presented and on the reliability of the underlying model (Rosenbacke et al., 2024; Aravindkumar et al., 2026).

Figure 5 summarizes how the choice of XAI technique tends to track the underlying data modality, and how both routes are aimed at the same ultimate goal of calibrated, appropriately-placed clinician trust rather than blind deference to the model.

The specific XAI method chosen tends to depend heavily on the underlying model and data type. For structured EHR data, SHAP and LIME dominate because they can attribute a prediction to individually meaningful variables such as a lab value or a vital sign trend, giving a clinician something directly comparable to the kind of reasoning they already use. For imaging, saliency-based approaches such as Grad-CAM are preferred because they produce a visual heat-map overlay on the scan itself, letting a radiologist see which region of the image drove the model's output, a format that maps naturally onto how radiologists already read images (Muhammad & Bendeche, 2024; Aravindkumar et al., 2026). Neither family of methods, however, fully resolves the underlying tension: an explanation can be faithful to what the model actually computed without being a reliable causal account of the true underlying biology, and reviews of the space increasingly call for standardized faithfulness and trustworthiness metrics so that an "explainable" model can be evaluated on the quality of its explanations, not simply the presence of one (Aravindkumar et al., 2026).

Across jurisdictions, regulatory approaches differ in detail, but recurring principles include risk-based classification, clinical validation, post-market monitoring, transparency, and human oversight. Designing audit trails and governance mechanisms from the beginning can make it easier for developers to adapt to changing regulatory requirements (Mohammed & Malhotra, 2025).

4.8 Future Trends and Research Directions

The future of AI in healthcare is shaped by several recurring trends in the literature, including multimodal integration, privacy-preserving learning, foundation models and generative AI, more mature explainability standards, continuous monitoring, equity-aware validation, human-AI collaboration, and applications

designed for low-resource settings (Zeng et al., 2026; Khan et al., 2026; Aravindkumar et al., 2026; Shaik et al., 2023).

Multi-modal integration. The field is transitioning from separate imaging, genomics, and EHR modeling problems to architectures that combine all three modalities, with the hope of obtaining a more comprehensive understanding of how interactions between a patient's molecular profile, imaging results, and longitudinal clinical history can be used to better predict the clinical course of disease (Khan et al., 2026; Zeng et al., 2026).

Federated and privacy-preserving learning. While institutions may not want to share sensitive information, they also want models that are trained on different populations that will generalize well, federated learning architectures that train collaboratively across sites and do not move raw patient data are likely to become common infrastructure rather than a research novelty (Khalid et al., 2023; Zeng et al., 2026).

Foundation models and generative AI. In drug discovery and other healthcare applications, large pretrained models are increasingly being adapted for clinical text processing, medical question answering, and candidate-molecule generation, while evaluation of biological validity and safety remains essential (Ferreira & Carneiro, 2025; Niazi, 2025).

Maturing explainability standards. Explainability is increasingly treated as a design and evaluation requirement rather than a post hoc addition. Recent reviews emphasize the need to evaluate explanation faithfulness, reliability, and trustworthiness instead of simply reporting that an explanation method was used (Aravindkumar et al., 2026; Rosenbacke et al., 2024).

Continuous and ambient monitoring. Remote patient monitoring is likely to expand beyond established cardiac and chronic-disease applications, with adaptive models that learn individual baselines while maintaining privacy and clinical safety (Shaik et al., 2023; Ciccarelli et al., 2025; Gautam et al., 2022).

Equity-aware model development. Given how consistently algorithmic bias appears in the literature as a concern, future research is likely to place growing emphasis on dataset diversity, fairness auditing, and subgroup performance reporting as standard components of model validation rather than optional add-ons (Zeng et al., 2026; Khan et al., 2026).

Human-AI collaboration models. Rather than framing AI as either a fully autonomous decision-maker or a passive tool, the more durable direction in the literature treats it as a collaborative partner whose recommendations are meant to be interrogated, questioned, and occasionally overridden by the clinician. Research into how best to structure that interaction (how much explanation to surface, when to prompt a second opinion, and how to present model uncertainty) is likely to become as active a research area as the underlying predictive algorithms themselves (Rosenbacke et al., 2024; Mohammed & Malhotra, 2025).

Global health applications. Much of the current evidence base comes from well-resourced health systems. In lower-resource settings, AI/ML tools may need to be lightweight, robust to infrastructure constraints, and carefully validated in the target population before clinical use (Mohammed & Malhotra, 2025; Zeng et al., 2026).

Combined, these directions point toward the next generation of AI in healthcare, one in which gains in predictive accuracy are complemented by the less visible but essential work of ensuring interoperability, accountability, equity, and clinical integration of such systems.

5. Conclusions

AI and machine learning are moving from experimental development toward practical use in healthcare. The evidence reviewed here includes imaging classifiers evaluated on chest X-ray and CT data during the COVID-19 pandemic, machine-learning cardiovascular risk models that discriminated better than conventional risk scores in a recent meta-analysis, generative-chemistry pipelines that propose new drug candidates, and wearable-based monitoring systems designed to detect early clinical deterioration, although the strength of validation differs substantially across these domains. Alongside the achievements, however, are equally tangible barriers: sensitive data needs to be safeguarded, black box models need to be made interpretable enough to be trusted by clinicians, computational and infrastructural gaps need to be filled so that the benefits are not just available in well-resourced institutions, and the ethical and legal frameworks under which algorithmic decisions are made in medicine are still being drafted. Whether the surrounding infrastructure for privacy-preserving learning, explainability standards, and adaptive regulation can be built out quickly enough to make the performance of the AI/ML models dependable, fair, and truly trusted at the bedside will likely determine the trajectory of the field in the next several years.

Abbreviations

AI

Artificial Intelligence

ML

Machine Learning

EHR

Electronic Health Record

XAI

Explainable Artificial Intelligence

AUC

Area Under the Receiver Operating Characteristic Curve

PPV

Positive Predictive Value

CNN

Convolutional Neural Network
RPM
Remote Patient Monitoring
FDA
U.S. Food and Drug Administration
SHAP
SHapley Additive exPlanations
LIME
Local Interpretable Model-agnostic Explanations
Grad-CAM
Gradient-weighted Class Activation Mapping.

Declarations

AI-usage declaration

During manuscript preparation, the authors used OpenAI ChatGPT for drafting assistance, language editing, paraphrasing, and structural revision, and Anthropic Claude for revision support, language editing, and reference checking during the response to editor comments. AI tools were not used to generate or fabricate primary research data, statistical results, or evidence. Figure 3 was redrawn using published source statistics, and all AI-assisted text, references, and figure inputs were reviewed and verified by the authors, who remain responsible for the final content.

Ethics approval and consent to participate

Not applicable. This article is a literature-based review and does not report research involving human participants, human data, human tissue, or animals.

Consent for publication

Not applicable.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

No specific funding was reported for this work.

Authors' contributions

ERK, MKS and KHS contributed to the conception, literature synthesis, organization, and preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

ERK is a Ph.D. Research Scholar in the Department of Computer Science, Atmiya University, Rajkot, Gujarat, India. MKS and KHS are Ph.D. Research Scholars in the Faculty of Science (Computer Science), Bhakta Kavi Narsinh Mehta University, Junagadh, Gujarat, India.

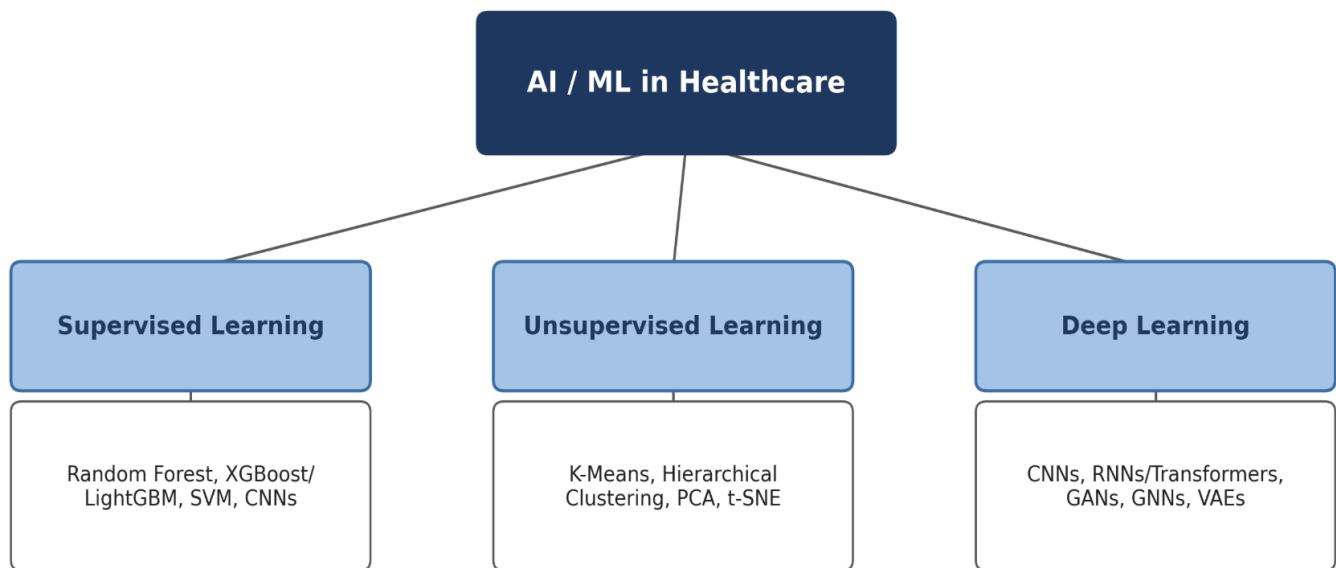
References

1. Alhumaidi, N. H., Dermawan, D., Kamaruzaman, H. F., & Alotaib, N. (2025). The use of machine learning for analyzing real-world data in disease prediction and management: Systematic review. *JMIR Medical Informatics*, 13, e68898. <https://doi.org/10.2196/68898>
2. Al-Juhani, A., Desoky, R., Iskander, Z., Alotaibi, R. M., Alzain, N. N., Aljohani, N., Alrefaai, M. M., Alhasanah, R. A., AlMehaimed, M. A., & Alharthi, A. S. (2025). Machine learning models for predicting mental health crises in adolescents using electronic health records: A systematic review. *Cureus*, 17(8), e89873. <https://doi.org/10.7759/cureus.89873>
3. Aravindkumar, A., Ramadoss, K., Fakhrudin Ahmed, S., Sampath, P., & Lakshminarayanan, V. (2026). Explainable AI in healthcare: A systematic review of XAI use cases in imaging, diagnostics, and rehabilitation. *Frontiers in Artificial Intelligence*, 9, 1749527. <https://doi.org/10.3389/frai.2026.1749527>
4. Askr, H., Elgeldawi, E., Aboul Ella, H., Elshaier, Y. A. M. M., Gomaa, M. M., & Hassanien, A. E. (2023). Deep learning in drug discovery: An integrative review and future challenges. *Artificial Intelligence Review*, 56, 5975–6037. <https://doi.org/10.1007/s10462-022-10306-1>
5. Bender, G., Faraj, S., & Bhardwaj, A. (2025). Implementation of AI in precision medicine. *arXiv*. <https://doi.org/10.48550/arXiv.2510.14194>
6. Berghout, T. (2025). The neural frontier of future medical imaging: A review of deep learning for brain tumor detection. *Journal of Imaging*, 11(1), 2. <https://doi.org/10.3390/jimaging11010002>
7. Bharati, S., Podder, P., Hossain Mondal, M. R., & Prasath, V. B. S. (2021). Medical imaging with deep learning for COVID-19 diagnosis: A comprehensive review. *arXiv*. <https://doi.org/10.48550/arXiv.2107.09602>
8. Ciccarella, M., Bramanti, A., Carrizzo, A., Garofano, M., Visco, V., Izzo, C., Rusciano, M. R., Galasso, G., Loria, F., Bruno, G., & Vecchione, C. (2025). Artificial intelligence-based remote monitoring for

- chronic heart failure: Design and rationale of the SMART-CARE study. *Frontiers in Digital Health*, 7, 1719562. <https://doi.org/10.3389/fdgth.2025.1719562>
9. Deimazar, G., & Sheikhtaheri, A. (2023). Machine learning models to detect and predict patient safety events using electronic health records: A systematic review. *International Journal of Medical Informatics*, 180, 105246. <https://doi.org/10.1016/j.ijmedinf.2023.105246>
 10. Dixon, D., Sattar, H., Moros, N., Kesireddy, S. R., Ahsan, H., Lakkimsetti, M., Fatima, M., Doshi, D., Sadhu, K., & Hassan, M. J. (2024). Unveiling the influence of AI predictive analytics on patient outcomes: A comprehensive narrative review. *Cureus*, 16(5), e59954. <https://doi.org/10.7759/cureus.59954>
 11. Dubey, P., & Tiwari, A. (2023). Artificial intelligence and remote patient monitoring in US healthcare market: A literature review. *Journal of Market Access & Health Policy*, 11(1), 2205618. <https://doi.org/10.1080/20016689.2023.2205618>
 12. Ferreira, F. J. N., & Carneiro, A. S. (2025). AI-driven drug discovery: A comprehensive review. *ACS Omega*. <https://doi.org/10.1021/acsomega.5c00549>
 13. Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., & Matic, A. (2022). Machine learning model to predict mental health crises from electronic health records. *Nature Medicine*, 28, 1240–1248. <https://doi.org/10.1038/s41591-022-01811-5>
 14. Gautam, N., Ghanta, S. N., Mueller, J., Mansour, M., Chen, Z., Puente, C., Ha, Y. M., Tarun, T., Dhar, G., Sivakumar, K., Zhang, Y., Halimeh, A. A., Nakarmi, U., Al-Kindi, S., DeMazumder, D., & Al'Aref, S. J. (2022). Artificial intelligence, wearables and remote monitoring for heart failure: Current and future applications. *Diagnostics*, 12(12), 2964. <https://doi.org/10.3390/diagnostics12122964>
 15. Hu, Q., Chen, Y., Zou, D., He, Y., & Xu, Y. (2024). Predicting adverse drug event using machine learning based on electronic health records: A systematic review and meta-analysis. *Frontiers in Pharmacology*, 15, 1497397. <https://doi.org/10.3389/fphar.2024.1497397>
 16. Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., & Lakshminarayanan, V. (2020). Deep learning based computer-aided systems for breast cancer imaging: A critical review. *Applied Sciences*, 10(22), 8298. <https://doi.org/10.3390/app10228298>
 17. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2, 573–584. <https://doi.org/10.1038/s42256-020-00236-4>
 18. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, 158, 106848. <https://doi.org/10.1016/j.compbiomed.2023.106848>
 19. Khan, S. N., Danishuddin, Khan, M. W. A., Guarnera, L., & Akhtar, S. M. F. (2026). Multi-modal AI in precision medicine: Integrating genomics, imaging, and EHR data for clinical insights. *Frontiers in Artificial Intelligence*, 8, 1743921. <https://doi.org/10.3389/frai.2025.1743921>
 20. Khare, S., Shaikh, M., & Havelikar, S. (2026). Artificial intelligence and precision medicine for optimizing patient care: A comprehensive review. *Intelligent Hospital*, 2(1), 100042. <https://doi.org/10.1016/j.inhs.2025.100042>

21. Koutroumpa, A., Papavasileiou, K., Papadiamantis, T., Melagraki, G., & Afantitis, A. (2023). A systematic review of deep learning methodologies used in the drug discovery process with emphasis on in vivo validation. *International Journal of Molecular Sciences*, 24(7), 6573. <https://doi.org/10.3390/ijms24076573>
22. Liu, T., Krentz, A., Lu, L., & Curcin, V. (2025). Machine learning based prediction models for cardiovascular disease risk using electronic health records data: Systematic review and meta-analysis. *European Heart Journal – Digital Health*, 6(1), 7–22. <https://doi.org/10.1093/ehjdh/ztae080>
23. Mohammed, R., & Malhotra, S. (2025). Ethical and regulatory challenges in machine learning-based healthcare systems: A review of implementation barriers and future directions. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 5(1), 100215. <https://doi.org/10.1016/j.tbench.2025.100215>
24. Muhammad, H., & Bendeche, M. (2024). Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis. *Computational and Structural Biotechnology Journal*, 24, 542–560. <https://doi.org/10.1016/j.csbj.2024.08.005>
25. Naik, K., Goyal, R. K., Foschini, L., Chak, C. W., Thielscher, C., Zhu, H., Lu, J., Lehár, J., Pacanowski, M. A., Terranova, N., Mehta, N., Korsbo, N., Fakhouri, T., Liu, Q., & Gobburu, J. (2024). Current status and future directions: The application of artificial intelligence/machine learning for precision medicine. *Clinical Pharmacology & Therapeutics*, 115(4), 673–686. <https://doi.org/10.1002/cpt.3152>
26. Niazi, S. K. (2025). Artificial intelligence in small-molecule drug discovery: A critical review of methods, applications, and real-world outcomes. *Pharmaceuticals*, 18(9), 1271. <https://doi.org/10.3390/ph18091271>
27. Rosenbacke, M., Melhus, Å., McKee, M., & Stuckler, D. (2024). How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: Systematic review. *JMIR AI*, 3, e53207. <https://doi.org/10.2196/53207>
28. Shaik, T., Tao, X., Higgins, N., Li, L., Gururajan, R., Zhou, X., & Acharya, U. R. (2023). Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *WIREs Data Mining and Knowledge Discovery*, 13(2), e1485. <https://doi.org/10.1002/widm.1485>
29. Singh, K., Ranapurwala, A., Bewoor, A., Patil, M., & Rai, S. (2023). Systematic review of techniques in brain image synthesis using deep learning. *arXiv*. <https://doi.org/10.48550/arXiv.2309.04511>
30. Sistaninejad, B., Rasi, H., & Nayeri, P. (2023). A review paper about deep learning for medical image analysis. *Computational and Mathematical Methods in Medicine*, 2023, 7091301. <https://doi.org/10.1155/2023/7091301>
31. Zeng, Q., Huang, C., & Zhu, J. (2026). AI-driven transformation of precision medicine: A comprehensive narrative review of key application areas, emerging paradigms, and future directions. *Frontiers in Public Health*, 13, 1656603. <https://doi.org/10.3389/fpubh.2025.1656603>

Figures



Model choice reflects a trade-off between predictive strength, robustness, and interpretability across data types (imaging, tabular EHR, sequential, molecular).

Figure 1

Representative algorithms for each of the three core AI/ML technique families that are commonly used in healthcare, and taxonomy thereof.

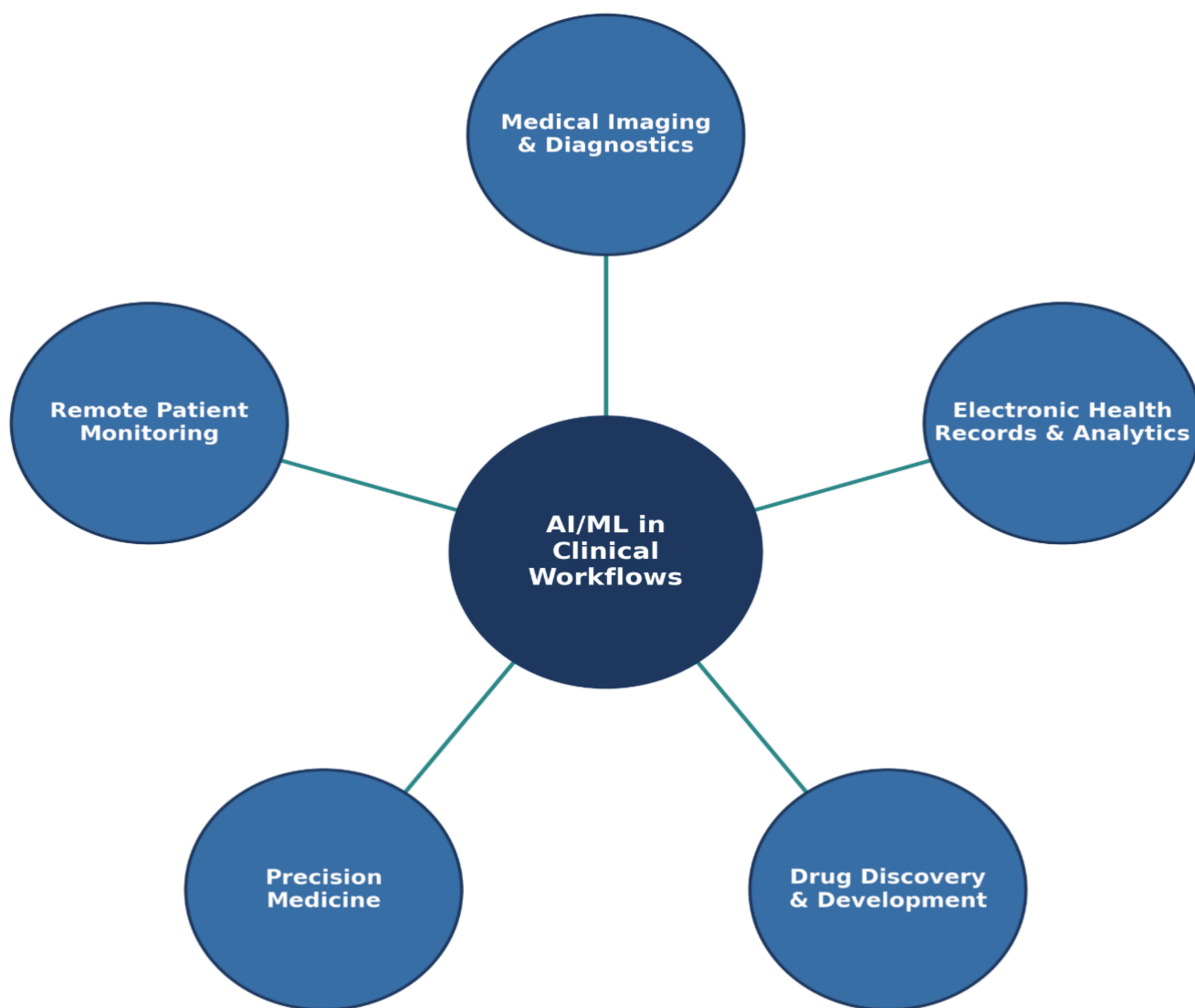


Figure 2

Five key areas of clinical or commercial impact that have been demonstrated with AI/ML

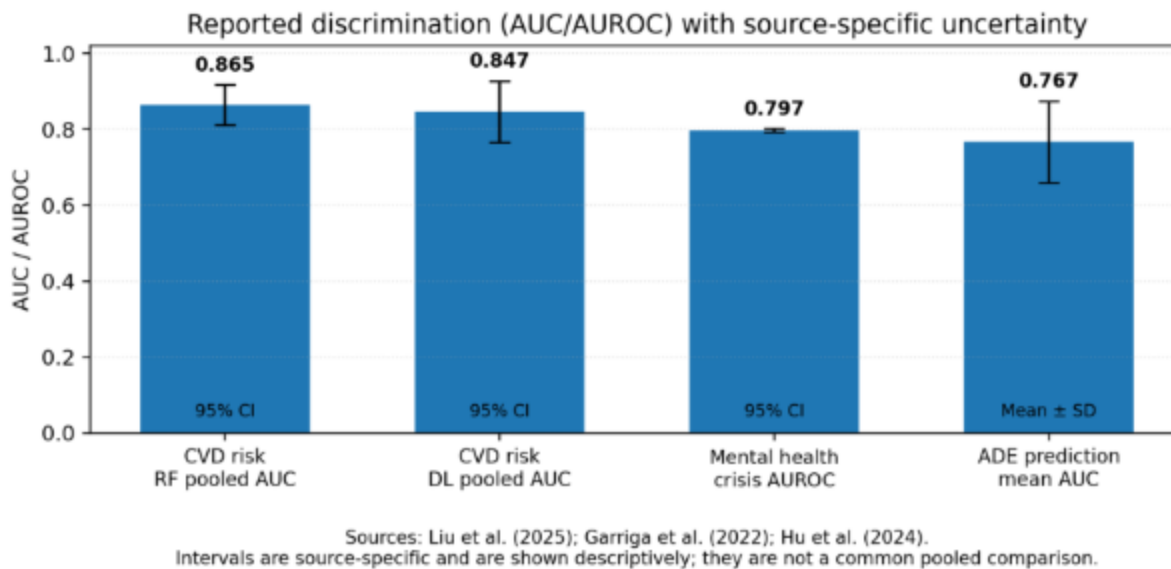


Figure 3

Reported discrimination values with source-specific uncertainty intervals. CVD risk values are pooled AUCs with 95% CIs (Liu et al., 2025); the mental-health crisis value is a single-study AUROC (patients aged 16 years and older in a general mental-health service) with its reported 95% CI (Garriga et al., 2022); and the adverse-drug-event value is the mean AUC $\pm$ SD across studies (Hu et al., 2024). The values are shown descriptively and should not be interpreted as a common pooled comparison.

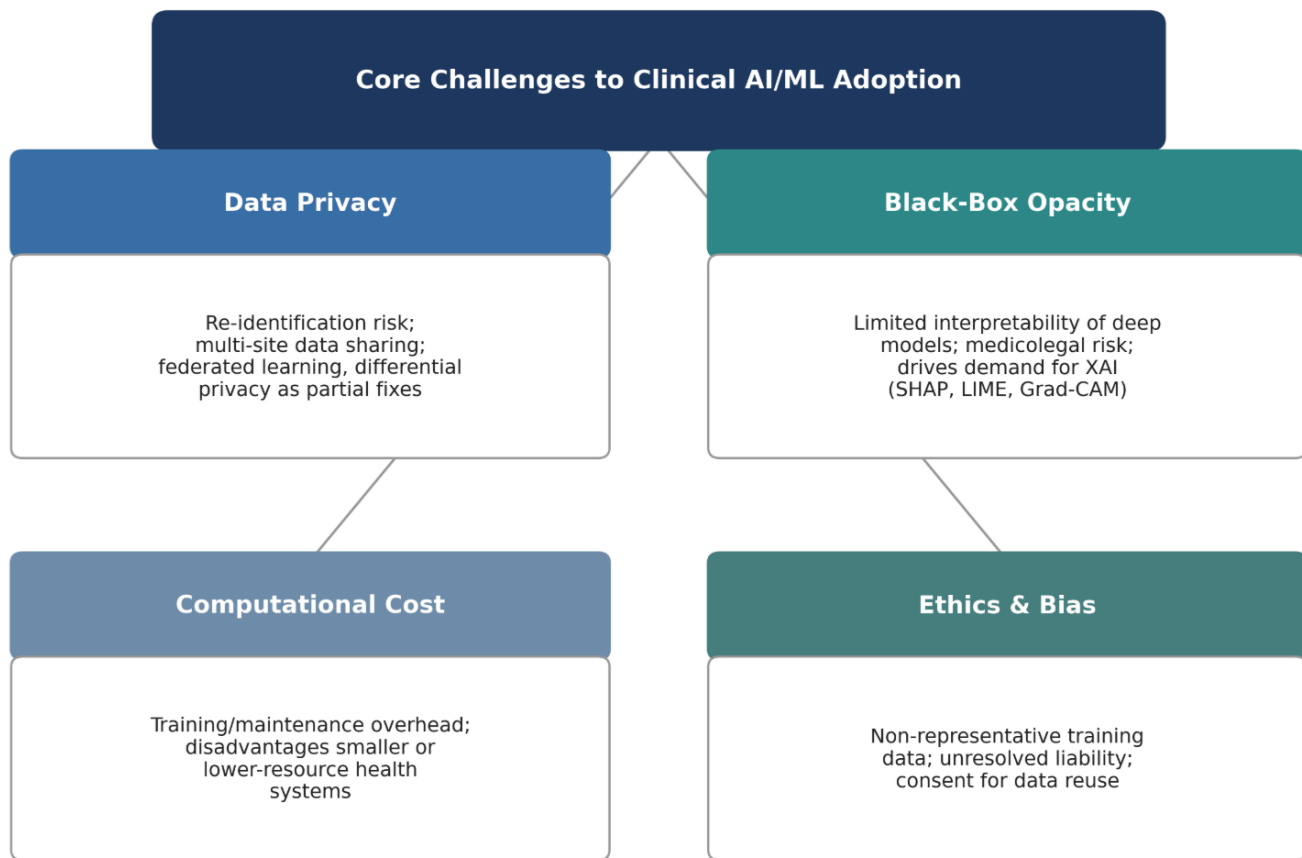


Figure 4

Four interlinked categories of challenge facing the clinical adoption of AI/ML in healthcare.

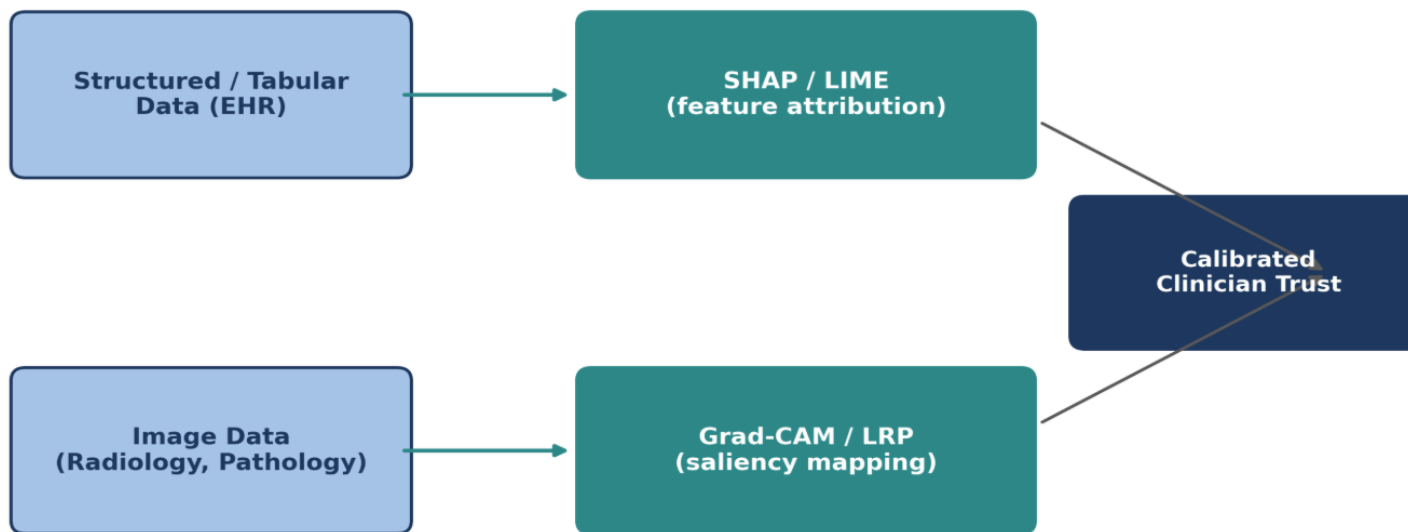


Figure 5

Mapping of explainable AI technique families to data modality, converging on calibrated clinician trust.