# Responsible adoption of multimodal artificial intelligence in health care: promises and challenges

*Ghazal Azarfar, Sara Naimimohasses, Sirisha Rambhatla, Matthieu Komorowski, Diana Ferro, Peter R Lewis, Darren Gates, Nawar Shara, Gregg M Gascon, Anthony Chang, Muhammad Mamdani, Mamatha Bhat, on behalf of the Alliance of Centers of Artificial Intelligence in Medicine working group*

**Clinicians rely on various data modalities—such as patient history, clinical signs, imaging, and laboratory results—to improve decision making. Multimodal artificial intelligence (AI) systems are emerging as powerful tools to process these diverse data types; however, the clinical adoption of multimodal AI systems is challenging because of data heterogeneity and integration complexities. The 2024 Temerty Centre for AI Research and Education in Medicine symposium, held on June 17, 2024, in Toronto, Canada, explored the potential and challenges of implementing multimodal AI in health care. In this Review, we summarise insights from the symposium. We discuss current applications, such as those used in early diagnosis of sepsis and cardiology, and identify key barriers, including fusion techniques, model selection, generalisation, fairness, safety, security, and international considerations on the responsible deployment of multimodal AI in health care. We outline practical strategies to overcome these obstacles, emphasising technologies such as federated learning to reduce bias and promote equitable health care. By addressing these challenges, multimodal AI can transform clinical practice and improve patient outcomes worldwide.**

*Lancet Digit Health* 2025; 7: 100917

Published Online December 11, 2025
https://doi.org/10.1016/j.landig.2025.100917

**Ajmera Transplant Centre** (G Azarfar PhD, Prof S Naimimohasses MD PhD, Prof M Bhat MD PhD) **and Toronto General Hospital Research Institute** (G Azarfar, Prof S Naimimohasses, Prof M Bhat), **University Health Network, Toronto, ON, Canada; Vector Institute, Toronto, ON, Canada** (G Azarfar, Prof M Mamdani PharmaD, Prof M Bhat)**; Department of Medicine** (Prof S Naimimohasses, Prof M Mamdani, Prof M Bhat), **Institute of Health Policy, Management and Evaluation** (Prof M Mamdani), **Dalla Lana School of Public Health** (Prof M Mamdani), **and Leslie Dan Faculty of Pharmacy** (Prof M Mamdani), **University of Toronto, Toronto, ON, Canada; Management and Sciences Engineering Department, University of Waterloo, Waterloo, ON, Canada** (Prof S Rambhatla PhD)**; Department of Surgery and Cancer, Imperial College London, London, UK** (Prof M Komorowski MD PhD)**; Department of Bioengineering, Imperial College London, London, UK** (Prof M Komorowski)**; Predictive and Preventive Medicine, Bambino Gesù Children's Hospital, IRCSS, Italy** (D Ferro PhD)**; Trustworthy AI Lab, Ontario Tech University, Oshawa, ON, Canada** (Prof P R Lewis PhD)**; Innovation Hub, Alder Hey Children's Hospital, Liverpool, England, UK** (D Gates MBChB)**; MedStar Health Research Institute, Georgetown University School-Medicine, Washington, DC, USA** (Prof N Shara PhD)**; OhioHealth, College of**

## Introduction

A 57-year-old man with lethargy and confusion presented to the emergency department. His medical history included a liver transplant 6 years ago for alcohol-related liver disease and subsequent graft cirrhosis. His family reported slurred speech, increasing fatigue, and abdominal distension.

On examination, the patient was drowsy with a Glasgow Coma Scale score of 12, afebrile, and mildly tachycardic (heart rate, 110 bpm) and had a blood pressure of 100/68 mm Hg. Although cardiorespiratory examination findings were normal, he had asterixis, and abdominal assessment revealed large-volume ascites. Abdominal x-ray confirmed ascites, but the finding was otherwise unremarkable. Laboratory findings showed normocytic anaemia with a haemoglobin level of 85 g/dL, mild leukopenia (white cell count, $3500 \times 10^6$/L), and elevated C-reactive protein level (20 mg/L). Renal and liver indices were stable. Diagnostic paracentesis revealed turbid fluid with a total white cell count of $1700 \times 10^6$/L, of which 60% were polymorphs. Treatment was started with albumin and broad-spectrum antimicrobials. Ascitic fluid culture results identified *Klebsiella* that caused spontaneous bacterial peritonitis, and subsequent blood cultures confirmed sepsis, exacerbating the pre-existing hepatic encephalopathy.

The case highlights the diagnostic complexity of sepsis in individuals who are immunocompromised, who often present with attenuated symptoms mimicking various non-infectious transplant-related complications.[1] Although sepsis is a leading cause of morbidity and mortality worldwide,[2] no single reliable biomarker exists,[3] requiring a comprehensive assessment of diverse data modalities for accurate diagnosis.

Artificial intelligence (AI) models have shown potential in sepsis diagnosis.[4] However, the clinical implementation of AI models has been challenging.[5] These models are often constrained by their reliance on tabular data extracted from patient reports, without incorporating time-sensitive, multimodal clinical data. A breakthrough might occur if AI can harness mathematical techniques to combine multiple data modalities and reveal their intricate relationships. Multimodal AI systems can operate continuously, integrating and analysing diverse data points that physicians cannot process cognitively.

The 2024 Temerty Centre for AI Research and Education in Medicine symposium[6] was held at the University of Toronto, Toronto, ON, Canada on June 17, 2024. The next day, an in-person Alliance of Centers of Artificial Intelligence in Medicine meeting attended by 50 participants was conducted. In preparation for the symposium's final discussion, two workshops, focusing on the potential of multimodal AI to improve clinical outcomes and the associated opportunities and challenges, were held in advance. This Review summarises key insights from the symposium (panel 1).

## What is multimodal AI?

The sepsis case illustrates how clinicians rely on multiple data modalities for decision making, with each modality providing unique insights into an individual's pathophysiological status. Multimodal AI refers to AI models capable of processing and integrating diverse data modalities, including text, image, tables, time series, video, and audio[7–9] (figure 1). The selection of modalities in a multimodal model depends on the complexity of the task; for instance, diagnosing acute respiratory distress syndrome requires both clinical and radiological data.[10]

Not all multimodal models are foundation models that are pre-trained on large datasets for broad adaptability. Nevertheless, foundation models are increasingly incorporating multimodal capabilities, particularly in vision–language tasks.[11,12] Vision–language tasks are especially

Medicine, The Ohio State University, Columbus, OH, USA (Prof G M Gascon PhD); Heart Failure Program, Children's Hospital of Orange County, Orange County, CA, USA (Prof A Chang MD); Data Science and Advanced Analytics, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada (Prof M Mamdani)

Correspondence to: Dr Mamatha Bhat, Department of Medicine, University of Toronto, Toronto, ON M5G 2N2, Canada **mamatha.bhat@uhn.ca**

*Panel 1:* **Summary of the discussion and findings of the 2024 Temerty Centre for AI Research and Education in Medicine (T-CAIREM) symposium**

**What is multimodal artificial intelligence (AI)?**

*Summary of discussion:* AI models that process multiple data modalities (eg, text, images, and tables) to improve accuracy.

*Concluding opinion*: Multimodal AI aligns well with health care, a field in which diverse data sources are essential.

**Why multimodal AI in health care?**

*Summary of discussion*: Multimodal AI has many applications in diagnostics, prognostics, and treatment. Moreover, models trained for a specific task can be repurposed across different organs, specialties, or modalities through transfer learning.

*Concluding opinion*: Multimodal AI can be the backbone of personalised medicine.

**Building a framework for multimodal AI in medicine**

*Summary of discussion*: There is no standardised benchmark for evaluating multimodal models. Fusion techniques such as early, joint, and late fusion were discussed.

*Concluding opinion*: A single-modality model can be designed as the baseline, and different fusion techniques can be attempted to select the best method.

**What are the challenges with multimodal AI?**

*Summary of discussion*: The challenges identified with multimodal AI were nature of multimodal data, model selection, validation, generalisation, interpretability, ease of use, safety, fairness, security, and international considerations.

*Concluding opinion*: We should ensure that measured data capture the phenomenon itself and not just its reflection. Addressing these challenges requires a cross-disciplinary and globally coordinated approach.

**Future perspectives**

*Summary of discussion*: Innovations in telemedicine, home monitoring, and the internet of medical things can democratise health-care access, especially in remote areas. Transparency, fairness, and accountability guidelines are required.

*Concluding opinion*: Current guidelines lack the necessary structure and detail to support innovation in health care.

relevant to health care owing to the field's reliance on medical imaging, documentation, and textual interpretation. This trend suggests the potential for developing multimodal foundation models in health care, representing a substantial advancement in AI maturity and precision health.[7–9,12–15]

## Why multimodal AI in health care?

Building on its capacity to improve clinical decision making, multimodal AI offers additional benefits. Multimodal AI can uncover hidden relationships among data modalities, mitigate physicians' cognitive blind spots, and expand health-care accessibility to remote regions. These capabilities extend AI's role beyond diagnostics, opening new avenues for personalised treatments and equitable health-care delivery.

Multimodal AI can also transform preventive care by combining demographics, social determinants, immune status, omics, imaging, and pharmacogenetics data. By analysing these modalities collectively, AI models can identify individuals who are at high risk of adverse outcome and might benefit from early intervention. For example, multimodal AI might predict who could benefit from colonoscopy screening before 50 years of age[16] or pinpoint subgroups for targeted vaccinations[17] or frequent monitoring.[18]

A multimodal AI system examines the cross-modal relationship between perceptual dimensions and stimulus attributes across different data domains. For example, the pitch and timing of adventitious sounds are informative for the differential diagnosis of respiratory infection versus heart failure. Multimodal AI can also compensate for missing modalities by leveraging other modalities depending on the design and clinical task. For example, in paediatric asthma management, where radiography is potentially harmful, multimodal AI can integrate data such as medical history, physical examination findings, and breath sounds to reduce the need for chest x-rays.[19]

In specialties such as cardiology, where various modalities are used, integrating modalities, including electrocardiogram, echocardiogram, CT, MRI, and nuclear scans, can be particularly valuable. Each data source provides unique insights into cardiac anatomy, function, and pathology. For example, electrocardiograms offer information on electrical activity, whereas echocardiograms help to visualise dynamic cardiac structure in real time. Electrocardiograms combined with echocardiograms enhance diagnostic accuracy beyond that by either of these modalities alone.[20]

Moreover, multimodal AI is valuable when patient–clinician communication is limited, particularly when dynamic, time-resolved models are used. For instance, in neonatal intensive care units, where decisions rely on diverse data such as maternal history, vital signs, cry recordings, electroencephalograms, and laboratory results,[21] multimodal AI can provide comprehensive patient assessments.

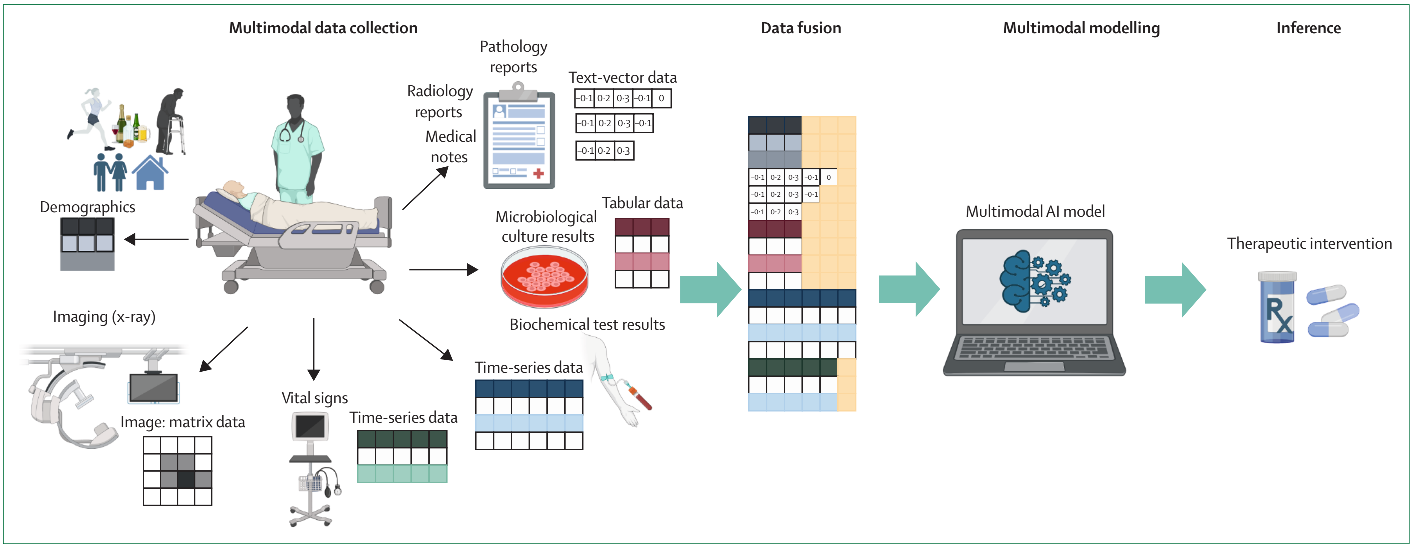


***Figure 1:*** **Schematic of multimodal examination**

Physicians can draw data from multiple modalities—such as tables (demographics), textual (medical notes, pathological reports, and radiological reports), time series (biochemical and microbiological test results and vital signs), and matrix data (x-ray images and digital pathology images)—to inform clinical decision making. Each modality is encoded into a vector representation and fused using an appropriate method. The multimodal AI model will analyse the processed multimodal data to support tailored therapeutic interventions. AI=artificial intelligence. Figure was created using BioRender.com and flaticon.com.

Future multimodal models could tailor treatments to individuals. For example, glioblastoma survival is influenced by demographic, genetic, and pharmacogenetic factors.[22] If multimodal AI can integrate these diverse data modalities, predictive patterns for optimal individualised treatment could be identified.

Another potential advantage of multimodal AI is cross-modal transfer learning, which enables models trained on one modality to be repurposed for another modality, especially when labelled datasets are scarce. For example, models trained on medical images and text need fewer labelled images than those trained on images alone for effective generalisation.[23] CT2US, a kidney ultrasound segmentation model, was initially trained on synthetic ultrasound images generated from CT scans and later fine-tuned and evaluated using real ultrasound images; the segmentation accuracy improved despite limited training dataset.[24] Cross-organ transfer of trained models has also been shown. For example, a model trained for multiple sclerosis segmentation using brain MRI scans was successfully fine-tuned for tumour classification using prostate multiparametric MRI scans.[25] Additionally, entity recognition, a natural language processing component that identifies textual categories—such as names, organisations, locations, expressions of times, quantities, and medical codes—can be repurposed across medical specialties using cross-specialty transfer learning with minimal annotation.[26]

These examples illustrate the potential applications of multimodal AI in health care. However, clinicians should understand that AI is complementary—critical thinking, ethical judgement, and therapeutic rapport remain the core of practice.

## Building a framework for multimodal AI in medicine

At the heart of multimodal AI is fusing heterogeneous data from diverse modalities to leverage their complementary aspects. Three common fusion strategies are used for combining multimodal data: early, joint, and late fusion.[8,9,15] Early fusion combines modalities after individual encoding. Early fusion is a potential starting point but might miss subtle cross-modal information.[27] Joint fusion encodes multiple modalities iteratively, with feedback linked to output loss, allowing feature optimisation during training. Although still evolving, joint fusion—especially transformer-based approaches—shows promise owing to its scalability and capacity to encode diverse modalities simultaneously. Late fusion involves training separate models for each modality and integrating their outputs through empirical aggregation methods such as averaging. However, late fusion might overlook correlations among input modalities, resulting in information loss.[15]

Developing a multimodal AI model starts by optimising single-modality models as baselines.[15] Once all data are available and single-modality modelling is completed, various fusion techniques can be compared to identify the best approach for a task. In addition, the baseline models can be retained for subsequent real-time deployment of multimodal systems to provide incremental predictions; this approach is particularly helpful in health care, a field in which data sources are commonly accessed at different time intervals—for example, 100 data points per second from a monitor, vital signs every 4 h, laboratory results every 12 h, and video recording of visits every 24 h. Deploying both baseline single-modality models and the multimodal model offers valuable insights for differential diagnosis and can inform knowledge-based modality weight adjustments for multimodal decision making.

The Holistic AI in Medicine framework is a multimodal AI system designed for a wide range of health-care applications.[8] This framework leverages early fusion for tabular, time-series, textual, and image data integration and advocates for using interpretable models such as XGBoost. The Holistic AI in Medicine framework showed consistent and robust performance, outperforming single-source models by 6–33% across various applications.[8] As we explore the potential of multimodal AI in health care, the challenges of its clinical deployment should be anticipated.

## What are the challenges with multimodal AI?

One key challenge with multimodal AI is managing large, heterogeneous multimodal datasets. A consistent methodology that includes ensuring consistent labelling across modalities and resolving discrepancies in diagnostic outputs is needed to harmonise these data. For example, laboratory results might not be consistent with radiology findings during the initial stages of a disease. Interactions across data modalities should be statistically and semantically examined across spatial and temporal dimensions.[28] In addition, variations in noise levels across modalities should be evaluated to reduce ambiguity. We refer readers to the study by Liang and colleagues[28] for a computer science perspective, whereas this Review provides the perspectives of health-care workers involved in the international deployment of AI tools.

Unlike physicians who rely on clinical judgement, AI models need standardised safeguards to maintain quality. The performance of the multimodal model relative to its single-modality counterpart models should justify the investments. The model's generalisation, interpretability, and compliance with patient rights should be ensured. Effective implementation requires robust international infrastructures and benchmarks (such as global guidelines, public multimodal datasets, and standardised evaluation metrics) for algorithm evaluation.[28] The National Academy of Medicine is developing an AI code of conduct to guide these efforts.[29] In the subsequent sections, we explore practical considerations and challenges in deploying multimodal AI in large scales (panel 2).

### Nature of multimodal data

Data access remains a substantial challenge to the development of multimodal AI systems in health care. Disparate information systems, heterogeneous data governance policies, and inconsistent data acquisition protocols hinder coherent multimodal dataset collection and integration. Nevertheless, progress is being made. Initiatives such as the Fast Healthcare Interoperability Resources[30] improve data exchange between different health-care systems. Advanced data harmonisation techniques, such as the Observational Medical Outcomes Partnership[31] standard data model, facilitate consistent data representation across institutions. Although these advancements gradually improve the accessibility and integrability of multimodal health-care data, capturing and analysing data in real time remains a hurdle.

As multimodal datasets such as Quilt-1M[32] become more accessible, a structured approach to data analysis is required in both research and clinical fields. The large size of these datasets poses challenges for storage, processing, and scalability. Labelling multimodal data, especially in health care, is labour intensive and requires domain expertise. In addition, multiple labels might be present across modalities, enabling exploration of label relationships and implementation of multitask learning, a strategy that has shown promise in research.[33]

The informational content of distinct data modalities can vary, displaying complementary or redundant characteristics, with different modalities holding varying degrees of relevance across tasks. For example, clinical notes describe symptoms and diagnosis, radiology images represent tissue density, and laboratory data reflect quantitative measurements. In addition, missingness across modalities might vary over time; this variation can be of different types such as missing at random (systematic differences between missing and observed values, described by variables such as sex or age), missing completely at random (similar distributions of missing and observed values),[34] missing not at random, and structurally missing data.

Additionally, aligning different modalities in multimodal datasets, particularly the alignment of time-series data with text, tabular, or imaging data, is challenging as collection dates might be missing.[28] Aligning such modalities requires maximising correlations and hidden relationships through temporal synchronisation or semantic coherence. Efficient multimodal data representation involves integrating and contextualising diverse modalities into a compressed format that maximises information density.

### Selecting the appropriate multimodal AI model

The selection of a multimodal AI model relies on understanding modality interactions and data structures. When prior knowledge is scarce, differentiable strategies can be leveraged to search for suitable model architecture in a completely data-driven manner. Initially, a candidate set of computational architectures is defined, and then, a meta-approach is used to identify the optimal architecture for a given task (figure 2).[38] In dynamic environments where AI models should continuously adapt to changes, reinforcement learning can be a viable approach. Reinforcement learning optimises decision making by maximising long-term rewards through short-term interactions with the environment.[39]

Medical phenomena are systems of interacting elements and can be described using multimodal graph representation learning. Many biological processes exhibit multimodal characteristics and hierarchical organisation, requiring higher-order reasoning in which abstract concepts depend on less abstract ones. For instance, although

**Panel 2: Challenges of deploying multimodal artificial intelligence (AI) in health care**

**Nature of multimodal data**
- *Example*: Disparate information systems, heterogeneous data governance policies, and inconsistent data acquisition protocols that hinder coherent multimodal data collection and integration.
- *Mitigation strategy*: Initiating data collection standards and using advanced data harmonisation techniques.

**Model selection**
- *Example*: Reliance on understanding modality interactions and data structures. However, detailed knowledge might not be available. For example, protein structures, their interactions with therapeutic compounds, and their associations with disease might be incompletely known.
- *Mitigation strategy*: Selecting a candidate set of models and using differentiable strategies to search for suitable model architecture in a completely data-driven manner.

**Validation**
- *Example*: Discordant information from different modalities. For example, during the early stages of rheumatoid arthritis, laboratory tests might show positive results for autoantibodies, but imaging might not reveal corresponding joint damages.
- *Mitigation strategy*: A collaborative dynamic between clinicians and software developers. Conventional metrics, along with novel approaches, are required. For example, weighed metrics can be incorporated to reflect the importance of modalities within context.

**Generalisation**
- *Example*: Difficulty in generalisation, even in one hospital, owing to population variations, disparities in health-care services, and local guidelines. The risk of model and data drift increases as each modality is introduced.
- *Mitigation strategy*: Collecting structured multimodal data, supporting research for generating synthetic multimodal data, selecting robust features resilient to input uncertainty, applying methods to enable decentralised training, and providing infrastructure for scalable AI architectures.

**Explainability and interpretability**
- *Example*: Black-box AI systems provide recommendations without sufficient explanation, challenging the obligation of physicians to explain the potential risks of a procedure to ensure informed consent.
- *Mitigation strategy*: Understanding AI decision making might not be necessary in medical practice, similar to how the functioning of some pharmaceutical treatments is not well understood. Where necessary, explainability and interpretability can be achieved by simplifying models, using feature engineering, and applying regularisation techniques.

**Ease of use**
- *Example*: Reduce instead of exacerbating clinical burden.
- *Mitigation strategy*: Design intuitive, user-friendly graphical user interfaces.

**Safety, liability, and continuous improvement**
- *Example*: Risks associated with minority groups, noisy data, or novel medical conditions.
- *Mitigation strategy*: Models should be accompanied by manual and thorough analyses of potential errors. Liability should be shared among software developers, medical staff, and health-care organisations.

**Fairness**
- *Example*: Children might not completely benefit from medical AI owing to challenges such as scarce paediatric-specific data, slower regulatory approvals, and smaller market than those for adults.
- *Mitigation strategy*: Understanding consequences of different forms of bias and addressing important ones during algorithmic design.

**Security and privacy**
- *Example*: Large language models trained on private hospital records might be manipulated through malicious prompts to spread misinformation about diseases or cause data breach.
- *Mitigation strategy*: Advanced technologies such as blockchain, federated learning, and homomorphic encryption.

**International considerations**
- *Example*: The EU has adopted a precautionary stance through its AI Act, focusing on transparency, accountability, and ethical safeguards. In contrast, the USA takes a more innovation-driven approach, guided by sector-specific guidelines and existing laws around civil rights, consumer protection, and medical device regulations.
- *Mitigation strategy*: Effective collaboration through organisations such as the UN.

detailed knowledge about protein structures, their interactions with therapeutic compounds, and their associations with diseases might be incomplete, general information about drug mechanisms is often available. In such cases, graph-based methods provide a practical modelling approach, in which concepts can be represented as nodes and relationships between them as edges in multimodal graphs.[40]

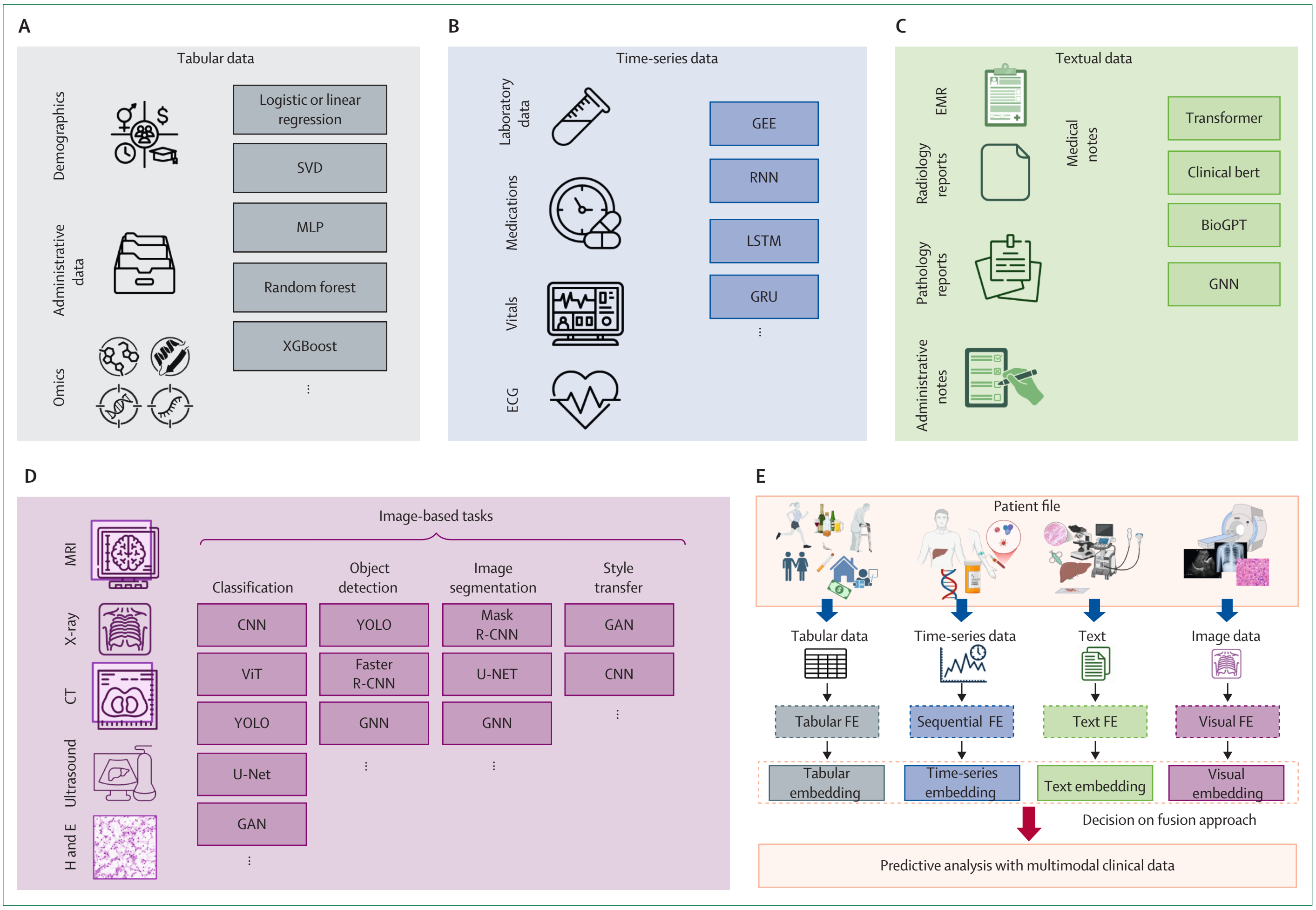


***Figure 2:*** **Model selection**

The choice of model architecture depends on the nature of the task and data modalities. (A) Models such as logistic or linear regression, SVD, MLP, Random Forest, and XGBoost are commonly used[34] for tabular data. (B) For time-series data with longitudinal relationships, methods such as GEE, RNN, LSTM, and GRU, are more appropriate.[35] (C) Transformers, GNNs, and pre-trained language models such as ClinicalBERT and BioGPT are effective[36] for text-based tasks. (D) For image-based tasks, vision-specific models such as CNN, ViT, YOLO, U-Net, GANs, and R-CNN are commonly used, depending on the application.[37] (E) In multimodal settings, data can be extracted from patient records across multiple modalities.[8] Each modality requires a suitable FE. The FE of each modality can be adopted from any of the models described in (A)–(D). Extracted features can then be fused using appropriate fusion strategies to enable predictive modelling of clinical outcomes. CNN=convolutional neural network. ECG=electrocardiogram. EMR=electronic medical record. FE=feature extractor. GAN=generative adversarial network. GEE=generalised estimating equations. GNN=graph neural network. GRU=gated recurrent unit. H&E=haematoxylin and eosin. LSTM=long short-term memory. MLP=multilayer perceptron. R-CNN=region-based convolutional neural network. RNN=recurrent neural network. SVD=singular value decomposition. ViT=vision transformer. YOLO=you only look once. Figure was created using BioRender.com and flaticon.com.

## Validation

The validation of multimodal AI models in health care requires both traditional and novel approaches to ensure robustness and clinical utility. Conventional metrics, such as area under the curve, accuracy, and mean squared error, remain applicable for supervised learning tasks. However, additional validation strategies are necessary when modalities provide discordant information. Many clinical symptoms might overlap across conditions. For example, in the early stages of rheumatoid arthritis, laboratory tests might show positive results for autoantibodies, whereas imaging findings might not reveal corresponding joint damages. Weighted metrics should be incorporated when certain modalities or specific timepoints are more important than others for a task. Validation should also address potential missing modalities and the timing of data availability (figure 3) to ensure that models remain reliable across different clinical scenarios.

Validation processes should integrate clinician-driven assessments, thus helping to determine the relative importance of each modality across various stages of a patient's treatment. For instance, in paediatric cardiology, the contribution of an echocardiogram might be greater than that of an MRI for a 1-year-old child. However, this balance might shift as the child grows and the quality of echocardiograms deteriorates. Furthermore, despite current preferences to certain modalities in clinical practice, AI might identify similar features in other modalities, making them more valuable than the current standards. For example, an AI model based on chest radiographs could identify signatures typically diagnosed using echocardiograms.[41]

Thus, a collaborative dynamic between clinicians and software developers can facilitate the analysis of medical questions and the relevance of data modalities.

## Generalisation and contextualisation

External validation of machine learning models is difficult owing to population differences, health-care disparities, local guidelines, cultural influences, and infrastructure limitations, which are particularly pronounced in resource-constrained settings.[42] Designers should consider diversity in data collection protocols, equipment, and software to ensure effective generalisation. Given the complexities in health care, generalisations within a particular hospital characterised by its unique workflow should be identified.

Even within a single hospital, the risk of model and data drift increases as each modality is introduced. The added input complexity might impose additional degrees of freedom, thereby causing performance degradation when data from sources other than those used for training are included. Temporal changes in dynamic clinical environments can also reduce model performance by causing dataset shifts, especially in advancing fields such as oncology or during unforeseen events such as pandemics.

Moreover, multimodal models trained on small datasets tend to over-rely on a single modality for decision making; hence, larger training datasets are necessary.[43] For instance, Khader and colleagues observed that their model initially focused on imaging data while disregarding clinical parameters. They implemented a vision-dropout technique, which forced the model to consider all available data modalities.[44] If the model relies more on a single modality, crafting task-specific architecture might introduce strong compositional biases, especially when diagnosing conditions with cross-modality similarities.

We believe that five key areas require attention to enhance model generalisation.

(1) Data collection: By implementing protocols for structured multimodal data collection[44]; (2) Data generation: By supporting research that generates synthetic multimodal data for model training or evaluation; (3) Feature selection: By selecting robust features resilient to input uncertainty using domain knowledge or machine learning embeddings and features representing static or dynamic modality-specific and modality-shared information;[45] 4) Model design: By designing models that can be fine-tuned using multimodality data from diverse locations and models that can be tailored to local needs.[46] Moreover, incorporating federated learning could enable decentralised training while ensuring data security;[47] (5) Scalable infrastructure: By providing infrastructure for scalable AI architectures capable of handling large multimodal datasets and parallelising computations across distributed resources.

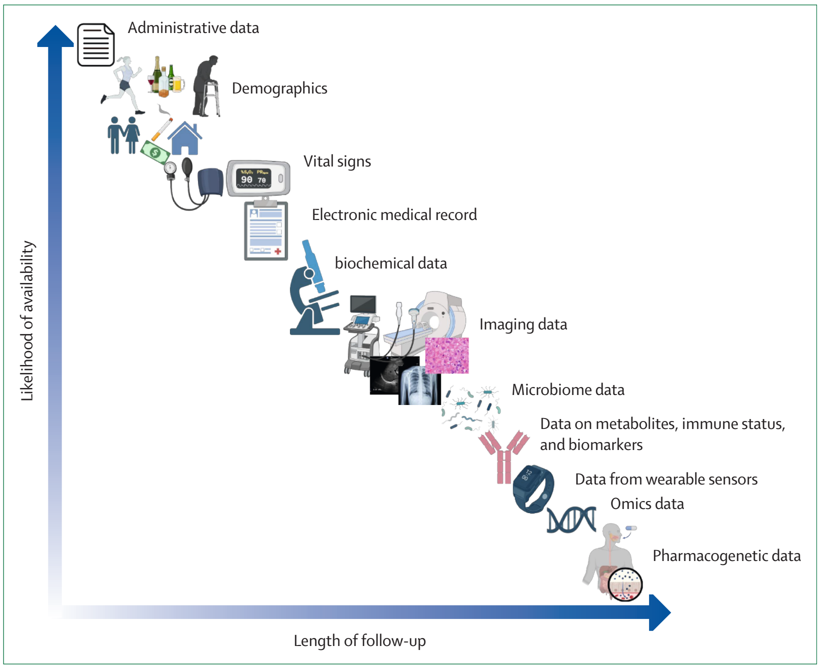


*Figure 3:* **Availability of clinical data modalities over time**
The availability of clinical data modalities varies considerably throughout a patient-care scenario. Most data are not accessible at the time of patient check-in but are acquired progressively during follow-up visits. Demographic and administrative information (eg, insurance status) is typically recorded shortly after check-in, whereas vital signs are collected during the initial nursing assessment. Diagnostic data, such as laboratory results and imaging, depend on physician orders and may only become available during or after subsequent visits. Less commonly collected modalities—such as a wearable sensors or omics data—are generally restricted to specific clinical contexts or research settings. Importantly, the timing and order in which these data become available will be shaped by the physician's diagnostic intuition, affecting the performance of the multimodal AI model. Figure was created using BioRender.com and flaticon.com.

## Explainability and interpretability

Another challenge is to design multimodal AI models that clinicians can understand. Medical practitioners are legally and ethically obligated to explain treatment risks to individuals, but black-box AI models complicate this responsibility. We believe that understanding AI decision making might not always be necessary in practice, similar to how some pharmaceutical treatments are not completely explainable by physicians owing to the inability to detail the exact biochemical mechanisms at play. Understanding model behaviour becomes crucial during early development and audit processes to acknowledge the validity of the underlying processes and identify potential biases.[48] In this context, the two AI model characteristics that should be considered are explainability and interpretability. An explainable AI model allows users to trace individual decisions but often requires a secondary post-hoc explanation model. An interpretable AI model is inherently understandable, eliminating the need for additional models to explain the decision-making process.[49] Notably, the secondary explanation models might introduce uncertainties; for instance, different explanation techniques, such as Shapley Additive Explanations[50] and Local Interpretable

Model-Agnostic Explanations,[51] might generate varying explanations for the same AI decision.

Explainability and interpretability can be achieved by simplifying models, using feature engineering, and applying regularisation techniques, such as selecting features by shrinking the coefficients of less important features, grounded in pre-existing knowledge. Recent advancements in multimodal AI focus on clarifying how each modality contributes to decision making[52] and demystifying black-box models through techniques such as feature attribution.[53] However, the complexity of deep neural networks in health care poses challenges to maintaining interpretability without compromising their performance. Interdisciplinary collaboration and continued clinician education are essential for developing AI models that are clinically relevant and trustworthy.

### Ease of use

AI can potentially reduce physician cognitive load by streamlining tasks, such as data collection, analysis, and decision making, and by facilitating asynchronous activities, such as patient communication. However, multimodal AI models might disrupt the clinical flow and increase physician load if not designed carefully. Furthermore, adopting multimodal AI technology in clinical settings depends on the perceived convenience of clinicians. User-friendly AI technologies that improve task efficiency by minimising errors and that require minimal learning by clinicians stand an increased chance of adoption in clinical practice. Additionally, if care delivery could be streamlined and compliance with standards could be ensured, regulatory approval could be obtained at an increased pace.[54,55]

### Safety, liability, and continuous improvement

AI models should be adopted into clinical practice only when they improve clinical prediction, outperform clinicians, or assist in managing complex cases. Multimodal AI models should perform better than their single-modality counterparts and be validated through pragmatic randomised controlled trials.

Addressing potential risks is crucial when integrating AI into clinical practice. Anomaly detection techniques can identify risks associated with minority groups, noisy data, or novel medical conditions. Models should come with a manual that includes thorough analyses of potential errors by considering clinical factors (eg, staff burnout or how a physician's diagnostic intuition affects the availability of various data modalities) and model limitations (eg, challenges in accurately predicting outcomes for outliers).[56] Errors might nevertheless occur despite such precautions, leading to complex liability issues.

Consequently, in cases of patient injury, it is often unclear who should be held responsible—the physician, health-care organisation, AI developer, or software manufacturer. However, to manage liability effectively, it is essential to determine whether the fault lies with the health-care provider, software designer, or manufacturer and to identify the nature of the error, whether it is data-related or systems-related.[57]

Most clinical AI models function as recommendation tools to support medical practice rather than provide definitive directives. Labelling the use of AI models as physician malpractice and placing liability solely on medical staff would increase the cognitive burden on medical staff and lead to resistance towards adopting AI tools. Conversely, assigning primary liability to software developers could obstruct innovation. We believe that the AI developer, the physician, the health system, and the regulators should share responsibility, with policy makers ensuring regulatory compliance during such scenarios.[58] Prioritising patient safety, thorough documentation, and investigation are also crucial in case of an error.

Continuous improvement requires educating the staff about AI system shortcomings, collaborating with developers to refine multimodal algorithms, and establishing monitoring mechanisms for ongoing performance evaluation.[59] Legal and ethical consultations are necessary to navigate potential litigation and ensure regulatory compliance. Regular reassessment of the AI system's risk management strategies is essential to maintain safety and reliability in clinical settings.

### Fairness

Bias is inherent in human cognition and is particularly evident when handling sensitive topics such as age, sex, gender, ethnicity, and socioeconomic status. In medicine, fairness is crucial for equitable distribution of resources and benefits across diverse demographics. However, some groups, such as children, might not completely benefit from new technologies, including medical AI, owing to challenges such as scarce paediatric-specific data, slower regulatory approvals, and a smaller market size than those for adults.[60]

The design of multimodal AI while ensuring fairness is challenging in aspects such as managing relationships and imbalances among data modalities that vary in scale or timepoints. The predictive nature of AI models implies that decisions are often made based on partial information and similarities to other cases, making it difficult to create an entirely unbiased model. The key challenge is understanding and mitigating the consequences of different forms of bias within a given context. Addressing the fairness challenge should involve a stakeholder-driven approach, in which quality measures for AI systems are co-designed by clinicians, engineers, people, and regulatory officers.[61]

For example, training AI models to achieve demographic parity by ensuring equal number of positive predictions for males and females might be suitable in contexts such as loan recommendations but inappropriate for breast cancer diagnosis. Clinicians and decision makers should critically evaluate AI systems and not rely solely on vendor claims of bias removal.

### Security and privacy

Data privacy is essential in health care because of legal and ethical implications, and breaches can have severe consequences such as affecting employment opportunities or increasing insurance premiums. Multimodal AI systems, which require extensive datasets, pose considerable privacy risks. Despite advances in differential privacy and anonymisation techniques, anonymised data can still be re-engineered to re-identify individuals,[62] posing considerable risks to patient confidentiality and highlighting the vulnerability of health-care data to adversarial attacks for extracting sensitive information. Concerns regarding the use of data by private AI companies and the absence of robust data governance frameworks also exist.[62]

The challenge of privacy management becomes increasingly crucial with the advent of multimodal large language models that interact directly with users. Often trained on private hospital records, these models can be manipulated through malicious prompts to spread misinformation about diseases or share private data.[63] Advanced technologies such as blockchain, federated learning, and homomorphic encryption offer potential solutions. Blockchain decentralises data storage to enhance security, federated learning reduces risks by not storing identifiable data within models, and homomorphic encryption allows models to be trained on encrypted datasets. However, these technologies require further development to be effectively implemented in clinical AI systems.[63]

### International considerations

Regulatory frameworks, clinical workflows, and quality of care vary considerably worldwide, especially between low-income and high-income countries. The EU and USA have distinct approaches to AI regulation, reflecting their different governance philosophies. The EU has adopted a precautionary stance through its AI Act, focusing on transparency, accountability, and ethical safeguards. In contrast, the USA takes a more innovation-driven approach, guided by sector-specific guidelines and existing laws around civil rights, consumer protection, and medical device regulations.

Successful implementation of multimodal AI systems requires broad international collaboration and leveraging local expertise to assess tailored risks. International governance should promote ethical development, thereby ensuring that AI standards are shaped by diverse perspectives and not solely by those from high-income countries. Data collection efforts should expand to include information from low-income and middle-income countries, with full respect for local data sovereignty, including authority over data sharing and using encrypted or trained models of low-income and middle-income countries. Such a coordinated approach fosters mutual benefit, supports equitable AI deployment, and reduces the risk of dominance by wealthy nations. WHO advocates inclusive governance, involving diverse stakeholders, such as the UN, financial institutions, regional organisations, civil society, and businesses, to work together for improved effectiveness in AI development.[64] Achieving universal health coverage with the aid of multimodal AI systems calls for a paradigm shift in multidisciplinary training, workforce development, and institutional processes to responsibly manage diverse data while preserving its legal, clinical, and ethical integrity.

## Future perspectives

A key advancement in multimodal AI is its ability to integrate and analyse diverse data streams in real time, including continuous data from wearable devices, home monitoring systems, electronic health records, and the internet of medical things. This comprehensive view of an individual's health enables rapid, informed decision making, potentially identifying early signs of deterioration, such as sepsis, and improving outcomes in critical care situations.[65]

Multimodal AI is expected to advance personalised medicine by integrating data from multiple omics (genomics, proteomics, and metabolomics) and clinical data to identify individualised risk factors. This integration allows for tailored prevention and treatment strategies based on each individual's genetic and molecular profile, and such an approach could be transformative in managing chronic diseases such as diabetes, cancer, and cardiovascular conditions, considerably improving patient outcomes and quality of life.[66–68] Additionally, advancements in pharmacogenomics might allow multimodal AI to predict patient responses to specific medications, thereby optimising drug efficacy and minimising adverse effects.[69]

Another potential of multimodal AI is bridging health-care gaps in underserved and remote regions. Multimodal AI can enhance the utility of telemedicine platforms to facilitate remote consultations and provide real-time diagnostic support, thus ensuring that individuals receive timely and accurate medical attention.[70] This democratisation of health care could lead to earlier disease diagnosis, improved chronic condition management, and overall enhanced health outcomes.

As multimodal AI becomes increasingly integrated into clinical practice, ethical and regulatory frameworks should evolve to address its unique challenges and opportunities. Future guidelines should ensure the transparency, fairness, and accountability of AI systems,[71] addressing issues such as algorithmic bias, informed consent, and interpretability of AI-driven decisions. Regulatory bodies might also establish standardised benchmarks for evaluating the performance and safety of multimodal AI systems.[72,73]

Another key future development in multimodal AI is the creation of systems capable of continuous learning and adaptation. Unlike static models, AI-based systems can be updated in real time as new data become available. Additionally, adaptive AI systems can personalise their learning to specific patient populations, thus improving their effectiveness in diverse clinical settings.[74] However, ensuring the safety of such systems presents substantial challenges,

## Search strategy and selection criteria

We searched the PubMed database for review, systematic review, and meta-analysis articles published between January, 2015, and May, 2024, to align with the period of increasing adoption of multimodal AI in health care. The search terms used were “multimodal data”, “deep learning”, “artificial intelligence”, “healthcare”, and “clinics”. A total of 597 records were retrieved. In addition, we performed multiple searches for grey literature in Google Scholar; for each search, the title and abstracts of top 100 relevant results were screened for eligibility. The Google Scholar searches were developed based on topics suggested and questions asked by the participants of the 2024 Temerty Centre for AI Research and Education in Medicine workshop. Only papers published in English were included, and studies that did not include health-related data were excluded. Bibliographies of shortlisted records were evaluated to ensure exhaustivity. The final reference list was generated based on originality and relevance to the broad scope of this Review.

and at the time of writing of this Review, no medical device equipped with continuous learning AI has been authorised for clinical use.

### Contributors

GA, AC, MM, and MB conceptualised the study. GA conducted the initial literature review and drafted the first version of the manuscript. SN contributed to writing of the clinical vignette. SR collected the literature and provided insights on multimodal artificial intelligence from a computer science perspective. MK drafted the future perspectives section. DF drafted the international considerations section and provided funding for her participation in the Alliance of Centers of Artificial Intelligence in Medicine meeting. PRL and NS drafted the fairness section. DG and GMG facilitated communications and provided feedback. All authors participated in the literature review, writing, and critical revision of the manuscript. MB provided the funding to support GA in the writing of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Declaration of interests

DF received support from the Italian Ministry of Health through the Current Research Fund and NextGenerationEU Piano Nazionale Complementare for attending the Alliance of Centers of Artificial Intelligence in Medicine meeting. MB received support from the University Health Network foundation (Transplant AI Initiative), Oncoustics, and Novo Nordisk for the current manuscript. All other authors including the members of the working group declare no competing interests.

### Acknowledgments

We acknowledge the support from the University Health Network Foundation (Transplant AI Initiative), Oncoustics, Novo Nordisk, and Paladin and thank the Italian Ministry of Health for supporting DF’s participation in this study through the Current Research Fund and NextGeneration EU PNC 0000001 D3 4 Health. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We also acknowledge the members of the Alliance of Centers of Artificial Intelligence in Medicine (ACAIM) who participated in the ACAIM workshop and shared their perspectives, which directly informed the development of this manuscript. We thank: Halim Abbas, Yin Aphinyanaphongs, James Barry, Helen Chen, Niloufar Eghbali Zarch, Khaled El Emam, Jake Franklin, Ioannis Kakadiaris, Diane Kim, Johanna Kim, Matthieu Komorowski, Jude Kong, April Khademi, Robert Kramer, Alexander Leichtle, Carson Leung, Alfonso Limon, Pietro Lio, Lisa Lix, Benoit L Marteau, Piyush Mathur, Paul Mullan, Jai Nahar, Micky Nnamdi, Anita Palepu, Dan Perri, Rubin Pillay, Sybil Russell, Suchi Saria, Ben Tamo, and May D Wang. During the preparation of this work, the authors used ChatGPT-4o to improve readability and language of the manuscript. After the initial draft and review by GA, each paragraph was individually copied and pasted into ChatGPT-4o with the following prompt: “Rewrite the following paragraph for publication in a high-impact AI in health-care journal such as *Lancet Digital Health*.” GA then reviewed and edited the result. After using the tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### References

1 Bafi AT, Tomotani DYV, De Freitas FG. Sepsis in solid-organ transplant patients. *Shock* 2017; **47:** 12–16.
2 Oczkowski S, Alshamsi F, Belley-Cote E, et al. Surviving Sepsis Campaign Guidelines 2021: highlights for the practicing clinician. *Pol Arch Intern Med* 2022; **132:** 16290.
3 Komorowski M, Green A, Tatham KC, Seymour C, Antcliffe D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine* 2022; **86:** 104394.
4 Deng HF, Sun MW, Wang Y, et al. Evaluating machine learning models for sepsis prediction: a systematic review of methodologies. *iScience* 2021; **25:** 103651.
5 Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021; **181:** 1065–70.
6 T-CAIREM 2024 Symposium. Multi-modal data and the future of health AI. June 17, 2024. https://tcairem.utoronto.ca/t-cairem-2024-symposium (accessed April 4, 2025).
7 Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med* 2022; **5:** 149.
8 Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. *NPJ Digit Med* 2022; **5:** 171.
9 Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med* 2022; **28:** 1773–84.
10 Matthay MA, Arabi Y, Arroliga AC, et al. A new global definition of acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2024; **209:** 37–47.
11 Zhang K, Zhou R, Adhikarla E, et al. A generalist vision–language foundation model for diverse biomedical tasks. *Nat Med* 2024; **30:** 3129–41.
12 Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature* 2023; **616:** 259–65.
13 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; **372:** 793–95.
14 Cai Q, Wang H, Li Z, Liu X. A survey on multimodal data-driven smart healthcare systems: approaches and applications. *IEEE Access* 2019; **7:** 133583–99.
15 Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med* 2020; **3:** 136.
16 Cancer Care Ontario. Colorectal cancer screening recommendations summary. https://www.cancercareontario.ca/en/guidelines-advice/cancer-continuum/screening/resources-healthcare-providers/colorectal-cancer-screening-summary (accessed July 21, 2024).
17 Azarfar G, Sun Y, Pasini E, et al. Using machine learning for personalized prediction of longitudinal coronavirus disease 2019 vaccine responses in transplant recipients. *Am J Transplant* 2025; **25:** 1107–16.
18 Iqbal U, Celi LA, Li YJ. How can artificial intelligence make medicine more preemptive? *J Med Internet Res* 2020; **22:** e17211.
19 Beyyumi E, Tawil MI, AlDhanhani H, et al. A single-institution experience in the use of chest radiographs for hospitalized children labeled as asthma exacerbation. *Front Pediatr* 2021; **9:** 722480.
20 Gupta-Malhotra M, Schaaf W, Kutty S. A primer on multimodal imaging and cardiology-radiology congenital heart interface. *Children (Basel)* 2019; **6:** 61.
21 Keles E, Bagci U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *NPJ Digit Med* 2023; **6:** 220.
22 Jovčevska I. Genetic secrets of long-term glioblastoma survivors. *Bosn J Basic Med Sci* 2019; **19:** 116–24.
23 Liang G, Greenwell C, Zhang Y, et al. Contrastive cross-modal pre-training: a general strategy for small sample medical imaging. *IEEE J Biomed Health Inform* 2022; **26:** 1640–49.

24 Song Y, Zheng J, Lei L, Ni Z, Zhao B, Hu Y. CT2US: cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. *Ultrasonics* 2022; **122:** 106706.

25 Lee J, Nishikawa RM. Cross-organ, cross-modality transfer learning: feasibility study for segmentation and classification. *IEEE Access* 2020; **8:** 210194–205.

26 Wang Z, Qu Y, Chen L, et al. Label-aware double transfer learning for cross-specialty medical named entity recognition. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018; 1–15.

27 Pawłowski M, Wróblewska A, Sysko-Romańczuk S. Effective techniques for multimodal data fusion: a comparative analysis. *Sensors (Basel)* 2023; **23:** 2381.

28 Liang PP, Zadeh A, Morency LP. Foundations and trends in multimodal machine learning: principles, challenges, and open questions. *ACM Comput Surv* 2024; **56:** 1–42.

29 Adams L, Fontaine E, Lin S, Crowell T, Chung VC, Gonzalez AA. Artificial intelligence in health, health care, and biomedical science: an AI code of conduct principles and commitments discussion draft. *NAM Perspect* 2024; published online April 8. https://doi.org/10.31478/202403a.

30 Vorisek CN, Lehne M, Klopfenstein SA, et al. Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: systematic review. *JMIR Med Inform* 2022; **10:** e35724.

31 Foundation for the National Institutes of Health. Observational medical outcomes partnership (OMOP). https://fnih.org/observational-medical-outcomes-partnership-omop (accessed Nov 19, 2025).

32 Ikezogwo WO, Seyfioglu MS, Ghezloo F, et al. Quilt-1M: one million image-text pairs for histopathology. *Adv Neural Inf Process Syst* 2023; **36:** 37995–8017.

33 Zadeh AB, Liang PP, Poria S, Cambria E, Morency LP. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018; 2236–46.

34 Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *Int J Epidemiol* 2014; **43:** 1336–39.

35 Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. *Inf Fusion* 2022; **81:** 84–90.

36 Cascarano A, Mur-Petit J, Hernandez-Gonzalez J, et al. Machine and deep learning for longitudinal biomedical data: a review of methods and applications. *Artif Intell Rev* 2023; **56** (suppl 2)**:** 1711–71.

37 Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: current research landscape and future directions. *J Am Med Inform Assoc* 2024; **31:** 1801–11.

38 Archana R, Jeevaraj PSE. Deep learning models for digital image processing: a review. *Artif Intell Rev* 2024; **57:** 11.

39 Shakya AK, Pillai G, Chakrabarty S. Reinforcement learning algorithms: a brief survey. *Expert Syst Appl* 2023; **231:** 120495.

40 Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. *Nat Biomed Eng* 2022; **6:** 1353–69.

41 Ueda D, Matsumoto T, Ehara S, et al. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. *Lancet Digit Health* 2023; **5:** e525–33.

42 Goetz L, Seedat N, Vandersluis R, van der Schaar M. Generalization—a key challenge for responsible AI in patient-facing clinical applications. *NPJ Digit Med* 2024; **7:** 126.

43 Eisermann A, Lee JH, Weber C, Wermter S. Generalization in multimodal language learning from simulation. Shenzhen, China: 2021 International Joint Conference on Neural Networks (IJCNN), 2021; 1–8.

44 Khader F, Kather JN, Müller-Franzes G, et al. Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data. *Sci Rep* 2023; **13:** 10666.

45 Dong H, Nejjar I, Sun H, Chatzi E, Fink O. SimMMDG: a simple and effective framework for multi-modal domain generalization. *Adv Neural Inf Process Syst* 2023; **36:** 78674–95.

46 Liang PP, Wu P, Ziyin L, Morency LP, Salakhutdinov R. Cross-modal generalization: learning in low resource modalities via Meta-Alignment. Proceedings of the 29th ACM International Conference on Multimedia, 2021; 2680–89.

47 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020; **3:** 119.

48 Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discov Artif Intell* 2024; **4:** 1–21.

49 Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019; **1:** 206–15.

50 Lundberg SM, Lee S. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), 2017; 4768–77.

51 Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 2016; 1135–44.

52 Tsai YH, Ma MQ, Yang M, Salakhutdinov R, Morency LP. Multimodal routing: improving local and global interpretability of multimodal language analysis. *Proc Conf Empir Methods Nat Lang Process* 2020; **2020:** 1823–33.

53 Merrick L, Taly A. The explanation game: explaining machine learning models using Shapley values. In: Holzinger A, Kieseberg P, Tjoa A, Weippl E, eds. Machine learning and knowledge extraction. Springer, 2020: 17–38.

54 Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int J Man Mach Stud* 1993; **38:** 475–87.

55 Gandhi TK, Classen D, Sinsky CA, et al. How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open* 2023; **6:** ooad079.

56 Mello MM, Guha N. Understanding liability risk from using health care artificial intelligence tools. *N Engl J Med* 2024; **390:** 271–78.

57 Holm S, Stanton C, Bartlett B. A new argument for no-fault compensation in health care: the introduction of artificial intelligence systems. *Health Care Anal* 2021; **29:** 171–88.

58 Maliha G, Gerke S, Cohen IG, Parikh RB. Artificial intelligence and liability in medicine: balancing safety and innovation. *Milbank Q* 2021; **99:** 629–47.

59 Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. *J Biomed Inform* 2020; **112:** 103611.

60 Ramgopal S, Sanchez-Pinto LN, Horvat CM, Carroll MS, Luo Y, Florin TA. Artificial intelligence-based clinical decision support in pediatrics. *Pediatr Res* 2023; **93:** 334–41.

61 Verma S, Rubin J. Fairness definitions explained. May 29, 2018. https://fairware.cs.umass.edu/papers/Verma.pdf (accessed Nov 20, 2025).

62 Williamson SM, Prybutok V. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Appl Sci* 2024; **14:** 675.

63 Rahman MA. A survey on security and privacy of multimodal LLMs–connected healthcare perspective. IEEE Globecom Workshops, 2023; 1807–12.

64 WHO. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. March 25, 2025. https://www.who.int/publications/i/item/9789240084759 (accessed June 20, 2025).

65 Ghiasi S, Zhu T, Lu P, et al. Sepsis mortality prediction using wearable monitoring in low–middle income countries. *Sensors (Basel)* 2022; **22:** 3866.

66 Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative multi-omics approaches in cancer research: from biological networks to clinical subtypes. *Mol Cells* 2021; **44:** 433–43.

67 Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol* 2021; **18:** 313–30.

68 Wang S, Yong H, He XD. Multi-omics: opportunities for research on mechanism of type 2 diabetes mellitus. *World J Diabetes* 2021; **12:** 1070–80.

69 Auwerx C, Sadler MC, Reymond A, Kutalik Z. From pharmacogenetics to pharmaco-omics: milestones and future directions. *HGG Adv* 2022; **3:** 100100.

70 Zhang X, Ma L, Sun D, Yi M, Wang Z. Artificial intelligence in telemedicine: a global perspective visualization analysis. *Telemed J E Health* 2024; **30:** e1909–22.

71 McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health* 2020; **2:** e221–23.

72 Larson DB, Harvey H, Rubin DL, Irani N, Tse JR, Langlotz CP. Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: summary and recommendations. *J Am Coll Radiol* 2021; **18:** 413–24.

73 Bengio Y, Mindermann S, Privitera D, et al. International scientific report on the safety of advanced AI: interim report. 2024. https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai/international-scientific-report-on-the-safety-of-advanced-ai-interim-report (accessed Aug 8, 2024).

74 US Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. 2024. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (accessed Aug 8, 2024).


© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).