



Multimodal biomedical AI

Julián N. Acosta¹, Guido J. Falcone¹, Pranav Rajpurkar^{2,4} and Eric J. Topol^{3,4}

The increasing availability of biomedical data from large biobanks, electronic health records, medical imaging, wearable and ambient biosensors, and the lower cost of genome and microbiome sequencing have set the stage for the development of multimodal artificial intelligence solutions that capture the complexity of human health and disease. In this Review, we outline the key applications enabled, along with the technical and analytical challenges. We explore opportunities in personalized medicine, digital clinical trials, remote monitoring and care, pandemic surveillance, digital twin technology and virtual health assistants. Further, we survey the data, modeling and privacy challenges that must be overcome to realize the full potential of multimodal artificial intelligence in health.

While artificial intelligence (AI) tools have transformed several domains (for example, language translation, speech recognition and natural image recognition), medicine has lagged behind. This is partly due to complexity and high dimensionality—in other words, a large number of unique features or signals contained in the data—leading to technical challenges in developing and validating solutions that generalize to diverse populations. However, there is now widespread use of wearable sensors and improved capabilities for data capture, aggregation and analysis, along with decreasing costs of genome sequencing and related ‘omics’ technologies. Collectively, this sets the foundation and need for novel tools that can meaningfully process this wealth of data from multiple sources, and provide value across biomedical discovery, diagnosis, prognosis, treatment and prevention.

Most of the current applications of AI in medicine have addressed narrowly defined tasks using one data modality, such as a computed tomography (CT) scan or retinal photograph. In contrast, clinicians process data from multiple sources and modalities when diagnosing, making prognostic evaluations and deciding on treatment plans. Furthermore, current AI assessments are typically one-off snapshots, based on a moment of time when the assessment is performed, and therefore not ‘seeing’ health as a continuous state. In theory, however, AI models should be able to use all data sources typically available to clinicians, and even those unavailable to most of them (for example, most clinicians do not have a deep understanding of genomic medicine). The development of multimodal AI models that incorporate data across modalities—including biosensors, genetic, epigenetic, proteomic, microbiome, metabolomic, imaging, text, clinical, social determinants and environmental data—is poised to partially bridge this gap and enable broad applications that include individualized medicine, integrated, real-time pandemic surveillance, digital clinical trials and virtual health coaches (Fig. 1). In this Review, we explore the opportunities for such multimodal datasets in healthcare; we then discuss the key challenges and promising strategies for overcoming these. Basic concepts in AI and machine learning will not be discussed here but are reviewed in detail elsewhere^{1–3}.

Opportunities for leveraging multimodal data

Personalized ‘omics’ for precision health. With the remarkable progress in sequencing over the past two decades, there has been a

revolution in the amount of fine-grained biological data that can be obtained using novel technical developments. These are collectively referred to as the ‘omes’, and includes the genome, proteome, transcriptome, immunome, epigenome, metabolome and microbiome⁴. These can be analyzed in bulk or at the single-cell level, which is relevant because many medical conditions such as cancer are quite heterogeneous at the tissue level, and much of biology shows cell and tissue specificity.

Each of the omics has shown value in different clinical and research settings individually. Genetic and molecular markers of malignant tumors have been integrated into clinical practice^{5,6}, with the US Food and Drug Administration (FDA) providing approval for several companion diagnostic devices and nucleic acid-based tests^{7,8}. As an example, Foundation Medicine and Oncotype IQ (Genomic Health) offer comprehensive genomic profiling tailored to the main classes of genomic alterations across a broad panel of genes, with the final goal of identifying potential therapeutic targets^{9,10}. Beyond these molecular markers, liquid biopsy samples—easily accessible biological fluids such as blood and urine—are becoming a widely used tool for analysis in precision oncology, with some tests based on circulating tumor cells and circulating tumor DNA already approved by the FDA¹¹. Beyond oncology, there has been a remarkable increase in the last 15 years in the availability and sharing of genetic data, which enabled genome-wide association studies¹² and characterization of the genetic architecture of complex human conditions and traits¹³. This has improved our understanding of biological pathways and produced tools such as polygenic risk scores¹⁴ (which capture the overall genetic propensity to complex traits for each individual), and may be useful for risk stratification and individualized treatment, as well as in clinical research to enrich the recruitment of participants most likely to benefit from interventions^{15,16}.

The integration of these very distinct types of data remains challenging. Yet, overcoming this problem is paramount, as the successful integration of omics data, in addition to other types such as electronic health record (EHR) and imaging data, is expected to increase our understanding of human health even further and allow for precise and individualized preventive, diagnostic and therapeutic strategies⁴. Several approaches have been proposed for multi-omics data integration in precision health contexts¹⁷. Graph neural networks are one example;^{18,19} these are deep learning model

¹Department of Neurology, Yale School of Medicine, New Haven, CT, USA. ²Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ³Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA. ⁴These authors jointly supervised this work: Pranav Rajpurkar, Eric J. Topol. ✉e-mail: pranav_rajpurkar@hms.harvard.edu; etopol@scripps.edu

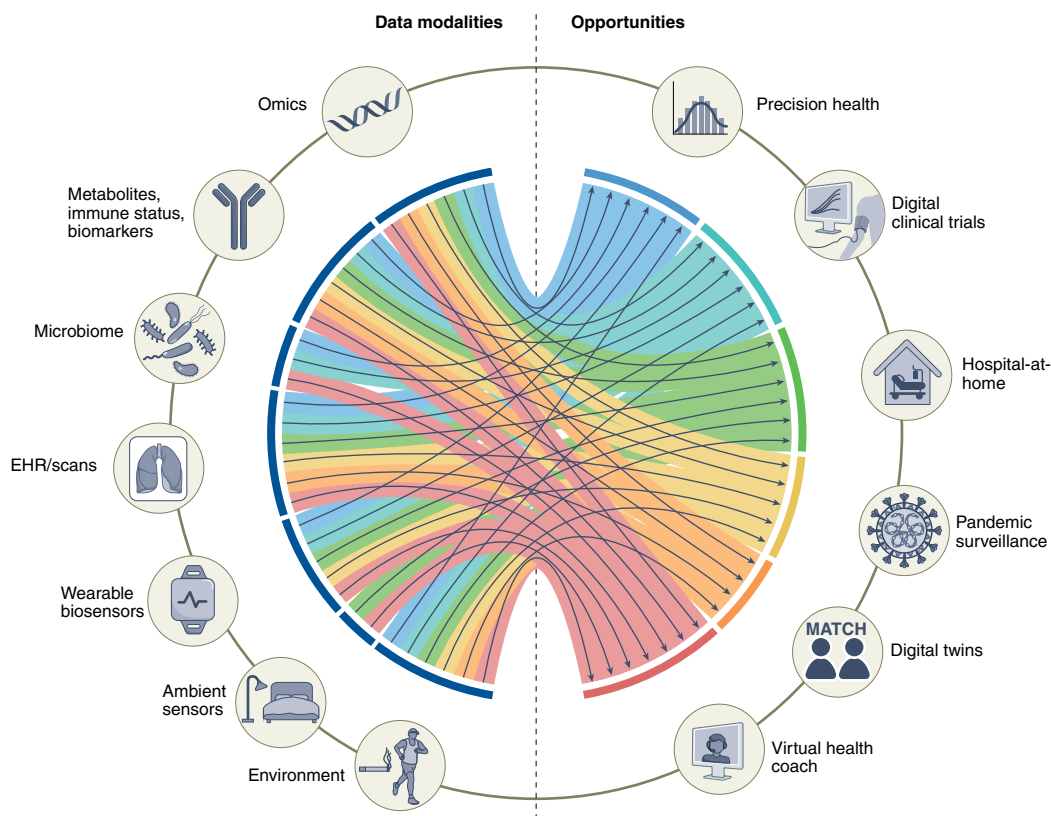


Fig. 1 | Data modalities and opportunities for multimodal biomedical AI. Created with BioRender.com.

architectures that process computational graphs—a well-known data structure comprising nodes (representing concepts or entities) and edges (representing connections or relationships between nodes)—thereby allowing scientists to account for the known inter-related structure of multiple types of omics data, which can improve performance of a model²⁰. Another approach is dimensionality reduction, including novel methods such as PHATE and Multiscale PHATE, which can learn abstract representations of biological and clinical data at different levels of granularity, and have been shown to predict clinical outcomes, for example, in people with coronavirus disease 2019 (COVID-19)^{21,22}.

In the context of cancer, overcoming challenges related to data access, sharing and accurate labeling could potentially lead to impactful tools that leverage the combination of personalized omics data with histopathology, imaging and clinical data to inform clinical trajectories and improve patient outcomes²³. The integration of histopathological morphology data with transcriptomics data, resulting in spatially resolved transcriptomics²⁴, constitutes a novel and promising methodological advancement that will enable finer-grained research into gene expression within a spatial context. Of note, researchers have utilized deep learning to leverage histopathology images to predict spatial gene expression from these images alone, pointing to morphological features in these images not captured by human experts that could potentially enhance the utility and lower the costs of this technology^{25,26}.

Genetic data are increasingly cost effective, requiring only a one-in-a-lifetime ascertainment, but they also have limited predictive ability on their own²⁷. Integrating genomics data with other omics data may capture more dynamic and real-time information on how each particular combination of genetic background and environmental exposures interact to produce the quantifiable continuum of health status. As an example, Kellogg et al.²⁸ conducted an *N*-of-1 study performing whole-genome sequencing (WGS) and

periodic measurements of other omics layers (transcriptome, proteome, metabolome, antibodies and clinical biomarkers); polygenic risk scoring showed an increased risk of type II diabetes mellitus, and comprehensive profiling of other omics enabled early detection and dissection of signaling network changes during the transition from health to disease.

As the scientific field advances, the cost-effectiveness profile of WGS will become increasingly favorable, facilitating the combination of clinical and biomarker data with already available genetic data to arrive at a rapid diagnosis of conditions that were previously difficult to detect²⁹. Ultimately, the capability to develop multimodal AI that includes many layers of omics data will get us to the desired goal of deep phenotyping of an individual; in other words, a true understanding of each person's biological uniqueness and how that affects health.

Digital clinical trials. Randomized clinical trials are the gold standard study design to investigate causation and provide evidence to support the use of novel diagnostic, prognostic and therapeutic interventions in clinical medicine. Unfortunately, planning and executing a high-quality clinical trial is not only time consuming (usually taking many years to recruit enough participants and follow them in time) but also financially very costly^{30,31}. In addition, geographic, sociocultural and economic disparities in access to enrollment, have led to a remarkable underrepresentation of several groups in these studies. This limits the generalizability of results and leads to a scenario whereby widespread underrepresentation in biomedical research further perpetuates existing disparities³². Digitizing clinical trials could provide an unprecedented opportunity to overcome these limitations, by reducing barriers to participant enrollment and retainment, promoting engagement and optimizing trial measurements and interventions. At the same time, the use of digital technologies can enhance the granularity of

the information obtained from participants, thereby increasing the value of these studies³³.

Data from wearable technology (including heart rate, sleep, physical activity, electrocardiography, oxygen saturation and glucose monitoring) and smartphone-enabled self-reported questionnaires can be useful for monitoring clinical trial patients, identifying adverse events or ascertaining trial outcomes³⁴. Additionally, a recent study highlighted the potential of data from wearable sensors to predict laboratory results³⁵. Consequently, the number of studies using digital products has been growing rapidly in the last few years, with a compound annual growth rate of around 34%³⁶. Most of these studies utilize data from a single wearable device. One pioneering trial used a 'band-aid' patch sensor for detecting atrial fibrillation; the sensor was mailed to participants who were enrolled remotely, without the use of any clinical sites, and set the foundation for digitized clinical trials³⁷. Many remote, site-less trials using wearables were conducted during the COVID-19 pandemic to detect coronavirus³⁸.

Effectively combining data from different wearable sensors with clinical data remains a challenge and an opportunity. Digital clinical trials could leverage multiple sources of participants' data to enable automatic phenotyping and subgrouping³⁴, which could be useful for adaptive clinical trial designs that use ongoing results to modify the trial in real time^{39,40}. In the future, we expect that the increased availability of these data and novel multimodal learning techniques will improve our capabilities in digital clinical trials. Of note, recent work in a time-series analysis by Google has demonstrated the promise of attention-based model architectures to combine both static and time-dependent inputs to achieve interpretable time-series forecasting. As a hypothetical example, these models could understand whether to focus on static features such as genetic background, known time-varying features such as time of the day or observed features such as current glycemic levels, to make predictions on future risk of hypoglycemia or hyperglycemia⁴¹. Graph neural networks have been recently proposed to overcome the problem of missing or irregularly sampled data from multiple health sensors, by leveraging information from the interconnection between these⁴².

Patient recruitment and retention in clinical trials are essential but remain a challenge. In this setting, there is an increasing interest in the utilization of synthetic control methods (that is, using external data to create controls). Although synthetic control trials are still relatively novel⁴³, the FDA has already approved medications based on historical controls⁴⁴ and has developed a framework for the utilization of real-world evidence⁴⁵. AI models utilizing data from different modalities can potentially help identify or generate the most optimal synthetic controls^{46,47}.

Remote monitoring: the 'hospital-at-home'. Recent progress with biosensors, continuous monitoring and analytics raises the possibility of simulating the hospital setting in a person's home. This offers the promise of marked reduction of cost, less requirement for healthcare workforce, avoidance of nosocomial infections and medical errors that occur in medical facilities, along with the comfort, convenience and emotional support of being with family members⁴⁸.

In this context, wearable sensors have a crucial role in remote patient monitoring. The availability of relatively affordable noninvasive devices (smartwatches or bands) that can accurately measure several physiological metrics is increasing rapidly^{49,50}. Combining these data with those derived from EHRs—using standards such as the Fast Healthcare Interoperability Resources, a global industry standard for exchanging healthcare data⁵¹—to query relevant information about a patient's underlying disease risk could create a more personalized remote monitoring experience for patients and caregivers. Ambient wireless sensors offer an additional opportunity to

collect valuable data. Ambient sensors are devices located within the environment (for example, a room, a wall or a mirror) ranging from video cameras and microphones to depth cameras and radio signals. These ambient sensors can potentially improve remote care systems at home and in healthcare institutions⁵².

The integration of data from these multiple modalities and sensors represents a promising opportunity to improve remote patient monitoring, and some studies have already demonstrated the potential of multimodal data in these scenarios. For example, the combination of ambient sensors (such as depth cameras and microphones) with wearables data (for example, accelerometers, which measure physical activity) has the potential to improve the reliability of fall detection systems while keeping a low false alarm rate⁵³, and to improve gait analysis performance⁵⁴. Early detection of impairments in physical functional status via activities of daily living such as bathing, dressing and eating is remarkably important to provide timely clinical care, and the utilization of multimodal data from wearable devices and ambient sensors can potentially help with accurate detection and classification of difficulties in these activities⁵⁵.

Beyond management of chronic or degenerative disorders, multimodal remote patient monitoring could also be useful in the setting of acute disease. A recent program conducted by the Mayo Clinic showcased the feasibility and safety of remote monitoring in people with COVID-19 (ref. ⁵⁶). Remote patient monitoring for hospital-at-home applications—not yet validated—requires randomized trials of multimodal AI-based remote monitoring versus hospital admission to show no impairment of safety. We need to be able to predict impending deterioration and have a system to intervene, and this has not been achieved yet.

Pandemic surveillance and outbreak detection. The current COVID-19 pandemic has highlighted the need for effective infectious disease surveillance at national and state levels⁵⁷, with some countries successfully integrating multimodal data from migration maps, mobile phone utilization and health delivery data to forecast the spread of the outbreak and identify potential cases^{58,59}.

One study has also demonstrated the utilization of resting heart rate and sleep minutes tracked using wearable devices to improve surveillance of influenza-like illness in the USA⁶⁰. This initial success evolved into the Digital Engagement and Tracking for Early Control and Treatment (DETECT) Health study, launched by the Scripps Research Translational Institute as an app-based research program aiming to analyze a diverse set of data from wearables to allow for rapid detection of the emergence of influenza, coronavirus and other fast-spreading viral illnesses. A follow-up study from this program showed that jointly considering participant self-reported symptoms and sensor metrics improved performance relative to either modality alone, reaching an area under the receiver operating curve value of 0.80 (95% confidence interval 0.73–0.86) for classifying COVID-19-positive versus COVID-19-negative status⁶¹.

Several other use cases for multimodal AI models in pandemic preparedness and response have been tested with promising results, but further validation and replication of these results are needed^{62,63}.

Digital twins. We currently rely on clinical trials as the best evidence to identify successful interventions. Interventions that help 10 of 100 people may be considered successful, but these are applied to the other 90 without proven or likely benefit. A complementary approach known as 'digital twins' can fill the knowledge gaps by leveraging large amounts of data to model and predict with high precision how a certain therapeutic intervention would benefit or harm a particular patient.

Digital twin technology is a concept borrowed from engineering that uses computational models of complex systems (for example, cities, airplanes or patients) to develop and test different strategies

or approaches more quickly and economically than in real-life scenarios⁶⁴. In healthcare, digital twins are a promising tool for drug target discovery^{65,66}.

Integrating data from multiple sources to develop digital twin models using AI tools has already been proposed in precision oncology and cardiovascular health^{67,68}. An open-source modular framework has also been proposed for the development of medical digital twin models⁶⁹. From a commercial point of view, Unlearn.AI has developed and tested digital twin models that leverage diverse sets of clinical data to enhance clinical trials for Alzheimer's disease and multiple sclerosis^{70,71}.

Considering the complexity of human organisms, the development of accurate and useful digital twin technology in medicine will depend on the ability to collect large and diverse multimodal data ranging from omics data and physiological sensors to clinical and sociodemographic data. This will likely require large collaborations across health systems, research groups and industry, such as the Swedish Digital Twins Consortium^{65,72}. The American Society of Clinical Oncology, through its subsidiary called CancerLinQ, developed a platform that enables researchers to utilize a wealth of data from patients with cancer to help guide optimal treatment and improve outcomes⁷³. The development of AI models capable of effectively learning from all these data modalities together, to make real-time predictions, is paramount.

Virtual health assistant. More than one-third of US consumers have acquired a smart speaker in the last few years. However, virtual health assistants—digital AI-enabled coaches that can advise people on their health needs—have not been developed widely to date, and those currently in the market often target a particular condition or use case. In addition, a recent review of health-focused conversational agents apps found that most of these rely on rule-based approaches and predefined app-led dialog⁷⁴.

One of the most popular, although not multimodal AI-based, current applications of these narrowly focused virtual health assistants is in diabetes care. Virta health, Accolade and Onduo by Verily (Alphabet) have all developed applications that aim to improve diabetes control, with some demonstrating improvement in hemoglobin A1c levels in individuals who followed the programs⁷⁵. Many of these companies have expanded or are in the process of expanding to other use cases such as hypertension control and weight loss. Other examples of virtual health coaches have tackled common conditions such as migraine, asthma and chronic obstructive pulmonary disease, among others⁷⁶. Unfortunately, most of these applications have been tested only on small observational studies, and much more research, including randomized clinical trials, are needed to evaluate their benefits.

Looking into the future, the successful integration of multiple data sources in AI models will facilitate the development of broadly focused personalized virtual health assistants⁷⁷. These virtual health assistants can leverage individualized profiles based on genome sequencing, other omics layers, continuous monitoring of blood biomarkers and metabolites, biosensors and other relevant biomedical data—to promote behavior change, answer health-related questions, triage symptoms or communicate with healthcare providers when appropriate. Importantly, these AI-enabled medical coaches will need to demonstrate beneficial effects on clinical outcomes via randomized trials to achieve widespread acceptance in the medical field. As most of these applications are focused on improving health choices, they will need to provide evidence of influencing health behavior, which represents the ultimate pathway for the successful translation of most interventions⁷⁸.

We still have a long way to go to achieve the full potential of AI and multimodal data integration into virtual health assistants, including the technical challenges, data-related challenges and privacy challenges discussed below. Given the rapid advances in

conversational AI⁷⁹, coupled with the development of increasingly sophisticated multimodal learning approaches, we expect future digital health applications to embrace the potential of AI to deliver accurate and personalized health coaching.

Multimodal data collection

The first requirement for the successful development of multimodal data-enabled applications is the collection, curation and harmonization of well-phenotyped and large annotated datasets, as no amount of technical sophistication can derive information not present in the data⁸⁰. In the last 20 years, many national and international studies have collected multimodal data with the ultimate goal of accelerating precision health (Table 1). In the UK, the UK Biobank initiated enrollment in 2006, reaching a final participant count of over 500,000, and plans to follow participants for at least 30 years after enrollment⁸¹. This large biobank has collected multiple layers of data from participants, including sociodemographic and lifestyle information, physical measurements, biological samples, 12-lead electrocardiograms and EHR data⁸². Further, almost all participants underwent genome-wide array genotyping and, more recently, proteome, whole-exome sequencing⁸³ and WGS⁸⁴. A subset of individuals also underwent brain magnetic resonance imaging (MRI), cardiac MRI, abdominal MRI, carotid ultrasound and dual-energy X-ray absorptiometry, including repeat imaging across at least two time points⁸⁵.

Similar initiatives have been conducted in other countries, such as the China Kadoorie Biobank⁸⁶ and Biobank Japan⁸⁷. In the USA, the Department of Veteran Affairs launched the Million Veteran Program⁸⁸ in 2011, aiming to enroll 1 million veterans to contribute to scientific discovery. Two important efforts funded by the National Institutes of Health (NIH) include the Trans-Omics for Precision Medicine (TOPMed) program and the All of Us Research Program. TOPMed collects WGS with the aim to integrate this genetic information with other omics data⁸⁹. The All of Us Research Program⁹⁰ constitutes another novel and ambitious initiative by the NIH that has enrolled about 400,000 diverse participants of the 1 million people planned across the USA, and is focused on enrolling individuals from broadly defined underrepresented groups in biomedical research, which is especially needed in medical AI^{91,92}.

Besides these large national initiatives, independent institutional and multi-institutional efforts are also building deep, multimodal data resources in smaller numbers of people. The Project Baseline Health Study, funded by Verily and managed in collaboration with Stanford University, Duke University and the California Health and Longevity Institute, aims to enroll at least 10,000 individuals, starting with an initial 2,500 participants from whom a broad range of multimodal data are collected, with the aim of evolving into a combined virtual-in-person research effort⁹³. As another example, the American Gut Project collects microbiome data from self-selected participants across several countries⁹⁴. These participants also complete surveys about general health status, disease history, lifestyle data and food frequency. The Medical Information Mart for Intensive Care (MIMIC) database⁹⁵, organized by the Massachusetts Institute of Technology, represents another example of multidimensional data collection and harmonization. Currently in its fourth version, MIMIC is an open-source database that contains de-identified data from thousands of patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center, including demographic information, EHR data (for example, diagnosis codes, medications ordered and administered, laboratory data and physiological data such as blood pressure or intracranial pressure values), imaging data (for example, chest radiographs)⁹⁶ and, in some versions, natural language text such as radiology reports and medical notes. This granularity of data is particularly useful for the data science and machine learning community, and MIMIC

Table 1 | Examples of studies with multimodal data available

Study	Country	Year started	Data modalities	Access	Sample size
UK Biobank	UK	2006	Questionnaires EHR/clinical Laboratory Genome-wide genotyping WES WGS Imaging Metabolites	Open access	~500,000
China Kadoorie Biobank	China	2004	Questionnaires Physical measurements Biosamples Genome-wide genotyping	Restricted access	~500,000
Biobank Japan	Japan	2003	Questionnaires Clinical Laboratory Genome-wide genotyping	Restricted access	~200,000
Million Veteran Program	USA	2011	EHR/clinical Laboratory Genome wide	Restricted access	1 million
TOPMed	USA	2014	Clinical WGS	Open access	~180,000
All of Us Research Program	USA	2017	Questionnaires SDH EHR/clinical Laboratory Genome wide Wearables	Open access	1 million (target)
Project Baseline Health Study	USA	2015	Questionnaires EHR/clinical Laboratory Wearables	Restricted access	10,000 (target)
American Gut Project	USA	2012	Clinical Diet Microbiome	Open access	~25,000
MIMIC	USA	2008–2019	Clinical/EHR Images	Open access	~380,000
MIPACT	USA	2018–2019	Wearables, clinical/EHR, physiological, laboratory	Restricted access	~6,000
North American Prodrome Longitudinal Study	USA	2008	Clinical Genetic	Restricted access	~1,000

SDH, social determinants of health; WES, whole-exome sequencing.

has become one of the benchmark datasets for AI models aiming to predict the development of clinical events such as kidney failure, or outcomes such as survival or readmissions^{97,98}.

The availability of multimodal data in these datasets may help achieve better diagnostic performance across a range of different tasks. As an example, recent work has demonstrated that the combination of imaging and EHR data outperforms each of these modalities alone to identify pulmonary embolism⁹⁹, and to differentiate between common causes of acute respiratory failure, such as heart failure, pneumonia or chronic obstructive pulmonary disease¹⁰⁰. The Michigan Predictive Activity & Clinical Trajectories in Health (MIPACT) study constitutes another example, with participants contributing data from wearables, physiological data (blood pressure), clinical information (EHR and surveys) and laboratory data¹⁰¹. The North American Prodrome Longitudinal Study is yet another example. This multisite program recruited individuals, and collected demographic, clinical and blood biomarker data with the goal of understanding the prodromal stages of psychosis^{102,103}. Other

studies focusing on psychiatric disorders such as the Personalised Prognostic Tools for Early Psychosis Management also collected several types of data and have already empowered the development of multimodal machine learning workflows¹⁰⁴.

Technical challenges

Implementation and modeling challenges. Health data are inherently multimodal. Our health status encompasses many domains (social, biological and environmental) that influence well-being in complex ways. Additionally, each of these domains is hierarchically organized, with data being abstracted from the big picture macro level (for example, disease presence or absence) to the in-depth micro level (for example, biomarkers, proteomics and genomics). Furthermore, current healthcare systems add to this multimodal approach by generating data in multiple ways: radiology and pathology images are, for example, paired with natural language data from their respective reports, while disease states are also documented in natural language and tabular data in the EHR.

Multimodal machine learning (also referred to as multimodal learning) is a subfield of machine learning that aims to develop and train models that can leverage multiple different types of data and learn to relate these multiple modalities or combine them, with the goal of improving prediction performance¹⁰⁵. A promising approach is to learn accurate representations that are similar for different modalities (for example, a picture of an apple should be represented similarly to the word 'apple'). In early 2021, OpenAI released an architecture termed Contrastive Language Image Pretraining (CLIP), which, when trained on millions of image–text pairs, matched the performance of competitive, fully supervised models without fine-tuning¹⁰⁶. CLIP was inspired by a similar approach developed in the medical imaging domain termed Contrastive Visual Representation Learning from Text (ConVIRT)¹⁰⁷. With ConVIRT, an image encoder and a text encoder are trained to generate image and text representations by maximizing the similarity of correctly paired image and text examples and minimizing the similarity of incorrectly paired examples—this is called contrastive learning. This approach for paired image–text co-learning has been used recently to learn from chest X-rays and their associated text reports, outperforming other self-supervised and fully supervised methods¹⁰⁸. Other architectures have also been developed to integrate multimodal data from images, audio and text, such as the Video-Audio-Text Transformer, which uses videos to obtain paired multimodal image, text and audio and to train accurate multimodal representations able to generalize with good performance on many tasks—such as recognizing actions in videos, classifying audio events, classifying images, and selecting the most adequate video for an input text¹⁰⁹.

Another desirable feature for multimodal learning frameworks is the ability to learn from different modalities without the need for different model architectures. Ideally, a unified multimodal model would incorporate different types of data (images, physiological sensor data and structured and unstructured text data, among others), codify concepts contained in these different types of data in a flexible and sparse way (that is, a unique task activates only a small part of the network, with the model learning which parts of the network should handle each unique task)¹¹⁰, produce aligned representations for similar concepts across modalities (for example, the picture of a dog, and the word 'dog' should produce similar internal representations), and provide any arbitrary type of output as required by the task¹¹¹.

In the last few years, there has been a transition from architectures with strong modality-specific biases—such as convolutional neural networks for images, or recurrent neural networks for text and physiological signals—to a relatively novel architecture called the Transformer, which has demonstrated good performance across a wide variety of input and output modalities and tasks¹¹². The key strategy behind transformers is to allow neural networks—which are artificial learning models that loosely mimic the behavior of the human brain—to dynamically pay attention to different parts of the input when processing and ultimately making decisions. Originally proposed for natural language processing, thus providing a way to capture the context of each word by attending to other words of the input sentence, this architecture has been successfully extended to other modalities¹¹³.

While each input token (that is, the smallest unit for processing) in natural language processing corresponds to a specific word, other modalities have generally used segments of images or video clips as tokens¹¹⁴. Transformer architectures allow us to unify the framework for learning across modalities but may still need modality-specific tokenization and encoding. A recent study by Meta AI (Meta Platforms) proposed a unified framework for self-supervised learning that is independent of the modality of interest, but still requires modality-specific preprocessing and training¹¹⁵. Benchmarks for self-supervised multimodal learning

allow us to measure the progress of methods across modalities: for instance, the Domain-Agnostic Benchmark for Self-supervised learning (DABS) is a recently proposed benchmark that includes chest X-rays, sensor data and natural image and text data¹¹⁶.

Recent advances proposed by DeepMind (Alphabet), including Perceiver¹¹⁷ and Perceiver IO¹¹⁸, propose a framework for learning across modalities with the same backbone architecture. Importantly, the input to the Perceiver architectures are modality-agnostic byte arrays, which are condensed through an attention bottleneck (that is, an architecture feature that restricts the flow of information, forcing models to condense the most relevant) to avoid size-dependent large memory costs (Fig. 2a). After processing these inputs, the Perceiver can then feed the representations to a final classification layer to obtain the probability of each output category, while the Perceiver IO can decode these representations directly into arbitrary outputs such as pixels, raw audio and classification labels, through a query vector that specifies the task of interest; for example, the model could output the predicted imaging appearance of an evolving brain tumor, in addition to the probability of successful treatment response.

A promising aspect of transformers is the ability to learn meaningful representations with unlabeled data, which is paramount in biomedical AI given the limited and expensive resources needed to obtain high-quality labels. Many of the approaches mentioned above require aligned data from different modalities (for example, image–text pairs). A study from DeepMind, in fact, suggested that curating higher-quality image–text datasets may be more important than generating large single-modality datasets, and other aspects of algorithm development and training¹¹⁹. However, these data may not be readily available in the setting of biomedical AI. One possible solution to this problem is to leverage available data from one modality to help learning with another—a multimodal learning task termed 'co-learning'¹⁰⁵. As an example, some studies suggest that transformers pretrained on unlabeled language data might be able to generalize well to a broad range of other tasks¹²⁰. In medicine, a model architecture called 'CycleGANs', trained on unpaired contrast and non-contrast CT scans, has been used to generate synthetic non-contrast or contrast CT scans¹²¹, with this approach showing improvements, for instance, in COVID-19 diagnosis¹²². While promising, this approach has not been tested widely in the biomedical setting and requires further exploration.

Another important modeling challenge relates to the exceedingly high number of dimensions contained in multimodal health data, collectively termed 'the curse of dimensionality'. As the number of dimensions (that is, variables or features contained in a dataset) increases, the number of people carrying some specific combinations of these features decreases (or for some combinations, even disappears), leading to 'dataset blind spots', that is, portions of the feature space (the set of all possible combinations of features or variables) that do not have any observation. These dataset blind spots can hurt model performance in terms of real-life prediction and should therefore be considered early in the model development and evaluation process¹²³. Several strategies can be used to mitigate this issue, and have been described in detail elsewhere¹²³. In brief, these include collecting data using maximum performance tasks (for example, rapid finger tapping for motor control, as opposed to passively collected data during everyday movement), ensuring large and diverse sample sizes (that is, with the conditions matching those expected at clinical deployment of the model), using domain knowledge to guide feature engineering and selection (with a focus on feature repeatability), appropriate model training and regularization, rigorous model validation and comprehensive model monitoring (including monitoring the difference between the distributions of training data and data found after deployment). Looking to the future, developing models able to incorporate previous knowledge (for example, known gene regulatory pathways and protein interactions)

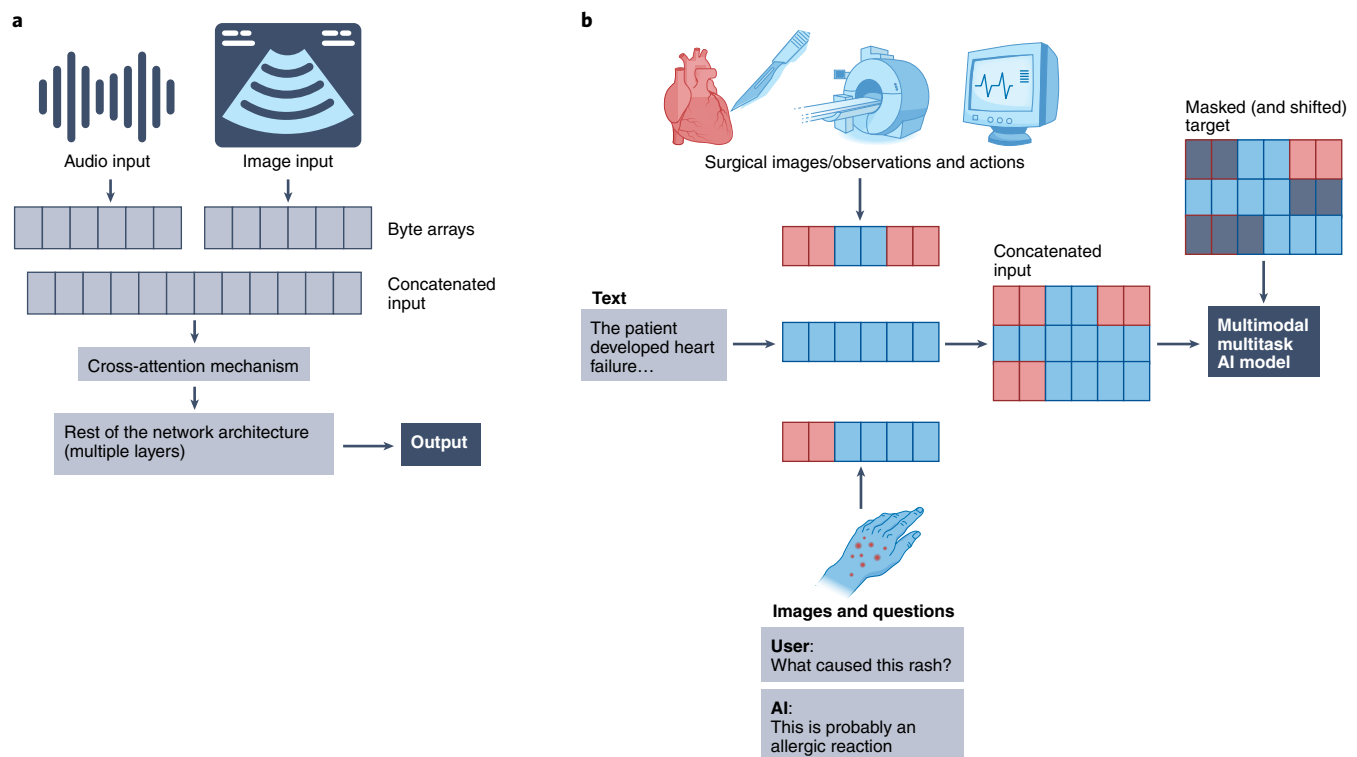


Fig. 2 | Simplified illustration of the novel technical concepts in multimodal AI. **a**, Simplified schematic of the Perceiver-like architecture: images, text and other inputs are converted agnostically into byte arrays that are concatenated (that is, fused) and passed through cross-attention mechanisms (that is, a mechanism to project or condense information into a fixed-dimensional representation) to feed information into the network. **b**, Simplified illustration of the conceptual framework behind the multimodal multitask architectures (for example, Gato), within a hypothetical medical example: distinct input modalities ranging from images, text and actions are tokenized and fed to the network as input sequences, with masked shifted versions of these sequences fed as targets (that is, the network only sees information from previous time points to predict future actions, only previous words to predict the next or only the image to predict text); the network then learns to handle multiple modalities and tasks.

might be another promising approach to overcome the curse of dimensionality. Along these lines, recent studies demonstrated that models augmented by retrieving information from large databases outperform larger models trained on larger datasets, effectively leveraging available information and also providing added benefits such as interpretability^{124,125}.

An increasingly used approach in multimodal learning is to combine the data from different modalities, as opposed to simply inputting several modalities separately into a model, to increase prediction performance—process termed ‘multimodal fusion’^{126,127}. Fusion of different data modalities can be performed at different stages of the process. The simplest approach involves concatenating input modalities or features before any processing (early fusion). While simple, this approach is not suitable for many complex data modalities. A more sophisticated approach is to combine and co-learn representations of these different modalities during the training process (joint fusion), allowing for modality-specific preprocessing while still capturing the interaction between data modalities. Finally, an alternative approach is to train separate models for each modality and combine the output probabilities (late fusion), a simple and robust approach, but at the cost of missing any information that could be abstracted from the interaction between modalities. Early work on fusion focused on allowing time-series models to leverage information from structured covariates for tasks such as forecasting osteoarthritis progression and predicting surgical outcomes in patients with cerebral palsy¹²⁸. As another example of fusion, a group from DeepMind used a high-dimensional EHR-based dataset comprising 620,000 dimensions that were projected into a continuous embedding space with only 800 dimensions,

capturing a wide array of information in a 6-h time frame for each patient, and built a recurrent neural network to predict acute kidney injury over time¹²⁹. A lot of studies have used fusion of two modalities (bimodal fusion) to improve predictive performance. Imaging and EHR-based data have been fused to improve detection of pulmonary embolism, outperforming single-modality models⁹⁹. Another bimodal study fused imaging features from chest X-rays with clinical covariates, improving the diagnosis of tuberculosis in individuals with HIV¹³⁰. Optical coherence tomography and infrared reflectance optic disc imaging have been combined to better predict visual field maps compared to using either of those modalities alone¹³¹.

Multimodal fusion is a general concept that can be tackled using any architectural choice. Although not biomedical, we can learn from some AI imaging work; modern guided image generation models such as DALL-E¹³² and GLIDE¹³³ often concatenate information from different modalities into the same encoder. This approach has demonstrated success in a recent study conducted by DeepMind (using Gato, a generalist agent) showing that concatenating a wide variety of tokens created from text, images and button presses, among others, can be used to teach a model to perform several distinct tasks ranging from captioning images and playing Atari games to stacking blocks with a robot arm (Fig. 2b)¹³⁴. Importantly, a recent study titled Align Before Fuse suggested that aligning representations across modalities before fusing them might result in better performance in downstream tasks, such as for creating text captions for images¹³⁵. A recent study from Google Research proposed using attention bottlenecks for multimodal fusion, thereby restricting the flow of cross-modality information to force models

to share the most relevant information across modalities and hence improving computational performance¹³⁶.

Another paradigm of using two modalities together is to ‘translate’ from one to the other. In many cases, one data modality may be strongly associated with clinical outcomes but be less affordable, accessible or require specialized equipment or invasive procedures. Deep learning-enabled computer vision has been shown to capture information typically requiring a higher-fidelity modality for human interpretation. As an example, one study developed a convolutional neural network that uses echocardiogram videos to predict laboratory values of interest such as cardiac biomarkers (troponin I and brain natriuretic peptide) and other commonly obtained biomarkers, and found that predictions from the model were accurate, with some of them even having more prognostic performance for heart failure admissions than conventional laboratory testing¹³⁷. Deep learning has also been widely studied in cancer pathology to make predictions beyond typical pathologist interpretation tasks with H&E stains, with several applications including prediction of genotype and gene expression, response to treatment and survival using only pathology images as inputs¹³⁸.

Many other important challenges relating to multimodal model architectures remain. For some modalities (for example, three-dimensional imaging), even models using only a single time point require large computing capabilities, and the prospect of implementing a model that also processes large-scale omics or text data represents an important infrastructural challenge.

While multimodal learning has improved at an accelerated rate for the past few years, we expect that current methods are unlikely to be sufficient to overcome all the major challenges mentioned above. Therefore, further innovation will be required to fully enable effective, multimodal AI models.

Data challenges. The multidimensional data underpinning health leads to a broad range of challenges in terms of collecting, linking and annotating these data. Medical datasets can be described along several axes¹³⁹, including the sample size, depth of phenotyping, the length and intervals of follow-up, the degree of interaction between participants, the heterogeneity and diversity of the participants, the level of standardization and harmonization of the data and the amount of linkage between data sources. While science and technology have advanced remarkably to facilitate data collection and phenotyping, there are inevitable trade-offs among these features of biomedical datasets. For example, although large sample sizes (in the range of hundreds of thousands to millions) are desirable in most cases for the training of AI models (especially multimodal AI models), the costs of achieving deep phenotyping and good longitudinal follow-up scales rapidly with larger numbers of participants, becoming financially unsustainable unless automated methods of data collection are put in place.

There are large-scale efforts to provide meaningful harmonization to biomedical datasets, such as the Observational Medical Outcomes Partnership Common Data Model developed by the Observational Health Data Sciences and Informatics collaboration¹⁴⁰. Harmonization enormously facilitates research efforts and enhances reproducibility and translation into clinical practice. However, harmonization may obscure some relevant pathophysiological processes underlying certain diseases. As an example, ischemic stroke subtypes tend not to be accurately captured by existing ontologies¹⁴¹, but utilizing raw data from EHRs or radiology reports could allow for the use of natural language processing for phenotyping¹⁴². Similarly, the Diagnostic and Statistical Manual of Mental Disorders categorizes diagnoses based on clinical manifestations, which might not fully represent underlying pathophysiological processes¹⁴³.

Achieving diversity across race/ethnicity, ancestry, income level, education level, healthcare access, age, disability status, geographic

locations, gender and sexual orientation has proven difficult in practice. Genomics research is a prominent example, with the vast majority of studies focusing on individuals from European ancestry¹⁴⁴. However, diversity of biomedical datasets is paramount as it constitutes the first step to ensure generalizability to the broader population¹⁴⁵. Beyond these considerations, a required step for multimodal AI is the appropriate linking of all data types available in the datasets, which represents another challenge owing to the increasing risk of identification of individuals and regulatory constraints¹⁴⁶.

Another frequent problem with biomedical data is the usually high proportion of missing data. While simply excluding patients with missing data before training is an option in some cases, selection bias can arise when other factors influence missing data¹⁴⁷, and it is often more appropriate to address these gaps with statistical tools, such as multiple imputation¹⁴⁸. As a result, imputation is a pervasive preprocessing step in many biomedical scientific fields, ranging from genomics to clinical data. Imputation has remarkably improved the statistical power of genome-wide association studies to identify novel genetic risk loci, and is facilitated by large reference datasets with deep genotypic coverage such as 1000 Genomes¹⁴⁹, the UK10K¹⁵⁰, the Haplotype reference consortium¹⁵¹ and, recently, TOPMed⁸⁹. Beyond genomics, imputation has also demonstrated utility for other types of medical data¹⁵². Different strategies have been suggested to make fewer assumptions. These include carry-forward imputation, with imputed values flagged and information added on when they were last measured¹⁵³, and more complex strategies such as capturing the presence of missing data and time intervals using learnable decay terms¹⁵⁴.

The risk of incurring several biases is important when conducting studies that collect health data, and multiple approaches are necessary to monitor and mitigate these biases¹⁵⁵. The risk of these biases is amplified when combining data from multiple sources, as the bias toward individuals more likely to consent to each data modality could be amplified when considering the intersection between these potentially biased populations. This complex and unsolved problem is more important in the setting of multimodal health data (compared to unimodal data) and would warrant its own in-depth review. Medical AI algorithms using demographic features such as race as inputs can learn to perpetuate historical human biases, thereby resulting in harm when deployed¹⁵⁶. Importantly, recent work has demonstrated that AI models can identify such features solely from imaging data, which highlights the need for deliberate efforts to detect racial bias and equalize racial outcomes during data quality control and model development¹⁵⁷. In particular, selection bias is a common type of bias in large biobank studies, and has been reported as a problem, for example, in the UK Biobank¹⁵⁸. This problem has also been pervasive in the scientific literature regarding COVID-19 (ref. 159). For example, patients using allergy medications were more likely to be tested for COVID-19, which leads to an artificially lower rate of positive tests, and an apparent protective effect among those tested—probably due to selection bias¹⁶⁰. Importantly, selection bias can result in AI models trained on a sample that differs considerably from the general population¹⁶¹, thus hurting these models at inference time¹⁶².

Privacy challenges. The successful development of multimodal AI in health requires breadth and depth of data, which encompasses higher privacy challenges than single-modality AI models. For example, previous studies have demonstrated that by utilizing only a little background information about participants, an adversary could re-identify those in large datasets (for example, the Netflix prize dataset), uncovering sensitive information about the individuals¹⁶³.

In the USA, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule is the fundamental legislation to protect privacy of health data. However, some types of health data—such as

user-generated and de-identified health data—are not covered by this regulation, which poses a risk of reidentification by combining information from multiple sources. In contrast, the more recent General Data Protection Regulation (GDPR) from the European Union has a much broader scope regarding the definition of health data, and even goes beyond data protection to also require the release of information about automated decision-making using these data¹⁶⁴.

Given the challenges, multiple technical solutions have been proposed and explored to ensure security and privacy while training multimodal AI models, including differential privacy, federated learning, homomorphic encryption and swarm learning^{165,166}. Differential privacy proposes a systematic random perturbation of the data with the ultimate goal of obscuring individual-level information while maintaining the global distribution of the dataset¹⁶⁷. As expected, this approach constitutes a trade-off between the level of privacy obtained and the expected performance of the models. Federated learning, on the other hand, allows several individuals or health systems to collectively train a model without transferring raw data. In this approach, a trusted central server distributes a model to each of the individuals/organizations; each individual or organization then trains the model for a certain number of iterations and shares the model updates back to the trusted central server¹⁶⁵. Finally, the trusted central server aggregates the model updates from all individuals/organizations and starts another round. Federated multimodal learning has been implemented in a multi-institutional collaboration for predicting clinical outcomes in people with COVID-19 (ref. ¹⁶⁸). Homomorphic encryption is a cryptographic technique that allows mathematical operations on encrypted input data, therefore providing the possibility of sharing model weights without leaking information¹⁶⁹. Finally, swarm learning is a relatively novel approach that, similarly to federated learning, is also based on several individuals or organizations training a model on local data, but does not require a trusted central server because it replaces it with the use of blockchain smart contracts¹⁷⁰.

Importantly, these approaches are often complementary and they can and should be used together. A recent study demonstrated the potential of coupling federated learning with homomorphic encryption to train a model to predict a COVID-19 diagnosis from chest CT scans, with the aggregate model outperforming all of the locally trained models¹²². While these methods are promising, multimodal health data are usually spread across several distinct organizations, ranging from healthcare institutions and academic centers to pharmaceutical companies. Therefore, the development of new methods to incentivize data sharing across sectors while preserving patient privacy is crucial.

An additional layer of safety can be obtained by leveraging novel developments in edge computing¹⁷¹. Edge computing, as opposed to cloud computing, refers to the idea of bringing computation closer to the sources of data (for example, close to ambient sensors or wearable devices). In combination with other methods such as federated learning, edge computing provides more security by avoiding the transmission of sensitive data to centralized servers. Furthermore, edge computing provides other benefits, such as reducing storage costs, latency and bandwidth usage. For example, some X-ray systems now run optimized versions of deep learning models directly in their hardware, instead of transferring images to cloud servers for identification of life-threatening conditions¹⁷².

As a result of the expanding healthcare AI market, biomedical data are increasingly valuable, leading to another challenge pertaining to data ownership. To date, this constitutes an open issue of debate. Some voices advocate for private patient ownership of the data, arguing that this approach would ensure the patients' right to self-determination, support health data transactions and maximize patients' benefit from data markets; while others suggest

Box 1 | Priorities for future development of multimodal biomedical AI

- Discover and formulate key medical AI tasks for which multimodal data will add value over single modalities.
- Develop approaches that can pretrain models using large amounts of unlabeled data across modalities and only require fine-tuning on limited labeled data.
- Benchmark the effect of model architectures and multimodal approaches when working with previously underexplored high-dimensional data, such as omics data.
- Collect paired (for example, image–text) multimodal data that could be used to train and test the generalizability of multimodal medical AI algorithms.

a non-property, regulatory model would better protect secure and transparent data use^{173,174}. Independent of the framework, appropriate incentives should be put in place to facilitate data sharing while ensuring security and privacy^{175,176}.

Conclusion

Multimodal medical AI unlocks key applications in healthcare and many other opportunities exist beyond those described here. The field of drug discovery is a pertinent example, with many tasks that could leverage multidimensional data including target identification and validation, prediction of drug interactions and prediction of side effects¹⁷⁷. While we addressed many important challenges to the use of multimodal AI, others that were outside the scope of this review are just as important, including the potential for false positives and how clinicians should interpret and explain the risks to patients.

With the ability to capture multidimensional biomedical data, we confront the challenge of deep phenotyping—understanding each individual's uniqueness. Collaboration across industries and sectors is needed to collect and link large and diverse multimodal health data (Box 1). Yet, as this juncture, we are far better at collating and storing such data, than we are at data analysis. To meaningfully process such high-dimensional data and actualize the many exciting use cases, it will take a concentrated joint effort of the medical community and AI researchers to build and validate new models, and ultimately demonstrate their utility to improve health outcomes.

Received: 21 March 2022; Accepted: 1 August 2022;

Published online: 15 September 2022

References

1. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29 (2019).
2. Esteva, A. et al. Deep learning-enabled medical computer vision. *NPJ Digit. Med.* **4**, 5 (2021).
3. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
4. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat. Rev. Genet.* **19**, 299–310 (2018).
5. Sidransky, D. Emerging molecular markers of cancer. *Nat. Rev. Cancer* **2**, 210–219 (2002).
6. Parsons, D. W. et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
7. Food and Drug Administration. List of cleared or approved companion diagnostic devices (in vitro and imaging tools) <https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools> (2021).
8. Food and Drug Administration. Nucleic acid-based tests <https://www.fda.gov/medical-devices/in-vitro-diagnostics/nucleic-acid-based-tests> (2020).
9. Foundation Medicine. Why comprehensive genomic profiling? <https://www.foundationmedicine.com/resource/why-comprehensive-genomic-profiling> (2018).

10. Oncotype IQ. Oncotype MAP pan-cancer tissue test <https://www.oncotypeiq.com/en-US/pan-cancer/healthcare-professionals/oncotype-map-pan-cancer-tissue-test/about-the-test-oncology> (2020).
11. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2018).
12. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Primers* **1**, 1–21 (2021).
13. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
14. Choi, S. W., Mak, T. S. -H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
15. Damask, A. et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. *Circulation* **141**, 624–636 (2020).
16. Marston, N. A. et al. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER trial. *Circulation* **141**, 616–623 (2020).
17. Duan, R. et al. Evaluation and comparison of multi-omics data integration methods for cancer subtyping. *PLoS Comput. Biol.* **17**, e1009224 (2021).
18. Kang, M., Ko, E. & Mersha, T. B. A roadmap for multi-omics data integration using deep learning. *Brief. Bioinform.* **23**, bbab454 (2022).
19. Wang, T. et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat. Commun.* **12**, 3445 (2021).
20. Zhang, X.-M., Liang, L., Liu, L. & Tang, M.-J. Graph neural networks and their current applications in bioinformatics. *Front. Genet.* **12**, 690049 (2021).
21. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
22. Kuchroo, M. et al. Multiscale PHATE identifies multimodal signatures of COVID-19. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01186-x> (2022).
23. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2021).
24. Marx, V. Method of the year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14 (2021).
25. He, B. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* **4**, 827–834 (2020).
26. Bergenstr hle, L. et al. Super-resolved spatial transcriptomics by deep data fusion. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01075-3> (2021).
27. Janssens, A. C. J. W. Validity of polygenic risk scores: are we measuring what we think we are? *Hum. Mol. Genet.* **28**, R143–R150 (2019).
28. Kellogg, R. A., Dunn, J. & Snyder, M. P. Personal omics for precision health. *Circ. Res.* **122**, 1169–1171 (2018).
29. Owen, M. J. et al. Rapid sequencing-based diagnosis of thiamine metabolism dysfunction syndrome. *N. Engl. J. Med.* **384**, 2159–2161 (2021).
30. Moore, T. J., Zhang, H., Anderson, G. & Alexander, G. C. Estimated costs of pivotal trials for novel therapeutic agents approved by the US food and drug administration, 2015–2016. *JAMA Intern. Med.* **178**, 1451–1457 (2018).
31. Sertkaya, A., Wong, H. -H., Jessup, A. & Belecche, T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin. Trials* **13**, 117–126 (2016).
32. Loree, J. M. et al. Disparity of race reporting and representation in clinical trials leading to cancer drug approvals from 2008 to 2018. *JAMA Oncol.* **5**, e191870 (2019).
33. Steinhubl, S. R., Wolff-Hughes, D. L., Nilsen, W., Iturriaga, E. & Califf, R. M. Digital clinical trials: creating a vision for the future. *NPJ Digit. Med.* **2**, 126 (2019).
34. Inan, O. T. et al. Digitizing clinical trials. *NPJ Digit. Med.* **3**, 101 (2020).
35. Dunn, J. et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nat. Med.* **27**, 1105–1112 (2021).
36. Marra, C., Chen, J. L., Coravos, A. & Stern, A. D. Quantifying the use of connected digital products in clinical research. *NPJ Digit. Med.* **3**, 50 (2020).
37. Steinhubl, S. R. et al. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. *JAMA* **320**, 146–155 (2018).
38. Pandit, J. A., Radin, J. M., Quer, G. & Topol, E. J. Smartphone apps in the COVID-19 pandemic. *Nat. Biotechnol.* **40**, 1013–1022 (2022).
39. Pallmann, P. et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med.* **16**, 29 (2018).
40. Klarin, D. & Natarajan, P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat. Rev. Cardiol.* <https://doi.org/10.1038/s41569-021-00638-w> (2021).
41. Lim, B., Arik, S.  , Loeff, N. & Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **37**, 1748–1764 (2021).
42. Zhang, X., Zeman, M., Tsiligrakis, T. & Zitnik, M. Graph-guided network for irregularly sampled multivariate time series. In *International Conference on Learning Representation (ICLR)*, (2022).
43. Thorlund, K., Dron, L., Park, J. J. H. & Mills, E. J. Synthetic and external controls in clinical trials—a primer for researchers. *Clin. Epidemiol.* **12**, 457–467 (2020).
44. Food and Drug Administration. FDA approves first treatment for a form of Batten disease <https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-form-batten-disease#:~:text=The%20U.S.%20Food%20and%20Drug,specific%20form%20of%20Batten%20disease> (2017).
45. Food and Drug Administration. Real-world evidence <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (2022).
46. AbbVie. Synthetic control arm: the end of placebos? <https://stories.abbvie.com/stories/synthetic-control-arm-end-placebos.htm> (2019).
47. Unlearn.AI. Generating synthetic control subjects using machine learning for clinical trials in Alzheimer's disease (DIA 2019) <https://www.unlearn.ai/post/generating-synthetic-control-subjects-alzheimers> (2019).
48. Noah, B. et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. *NPJ Digit. Med.* **1**, 20172 (2018).
49. Strain, T. et al. Wearable-device-measured physical activity and future health risk. *Nat. Med.* **26**, 1385–1391 (2020).
50. Iqbal, S. M. A., Mahgoub, I., Du, E., Leavitt, M. A. & Asghar, W. Advances in healthcare wearable devices. *NPJ Flex. Electron.* **5**, 9 (2021).
51. Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S. & Ramoni, R. B. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J. Am. Med. Inform. Assoc.* **23**, 899–908 (2016).
52. Haque, A., Milstein, A. & Fei-Fei, L. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature* **585**, 193–202 (2020).
53. Kwolek, B. & Kepski, M. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Comput. Methods Prog. Biomed.* **117**, 489–501 (2014).
54. Wang, C. et al. Multimodal gait analysis based on wearable inertial and microphone sensors. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)* 1–8 (2017).
55. Luo, Z. et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. In *Proc. Machine Learning Research* Vol. 85, 1–18 (PMLR, 2018).
56. Coffey, J. D. et al. Implementation of a multisite, interdisciplinary remote patient monitoring program for ambulatory management of patients with COVID-19. *NPJ Digit. Med.* **4**, 123 (2021).
57. Whitelaw, S., Mamas, M. A., Topol, E. & Van Spall, H. G. C. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit. Health* **2**, e435–e440 (2020).
58. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
59. Jason Wang, C., Ng, C. Y. & Brook, R. H. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *JAMA* **323**, 1341–1342 (2020).
60. Radin, J. M., Wineinger, N. E., Topol, E. J. & Steinhubl, S. R. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. *Lancet Digit. Health* **2**, e85–e93 (2020).
61. Quer, G. et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat. Med.* **27**, 73–77 (2020).
62. Syrowatka, A. et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digit. Med.* **4**, 96 (2021).
63. Varghese, E. B. & Thampi, S. M. A multimodal deep fusion graph framework to detect social distancing violations and FCGs in pandemic surveillance. *Eng. Appl. Artif. Intell.* **103**, 104305 (2021).
64. San, O. The digital twin revolution. *Nat. Comput. Sci.* **1**, 307–308 (2021).
65. Bj rnsson, B. et al. Digital twins to personalize medicine. *Genome Med.* **12**, 4 (2019).
66. Kamel Boulos, M. N. & Zhang, P. Digital twins: from personalised medicine to precision public health. *J. Pers. Med.* **11**, 745 (2021).
67. Hernandez-Boussard, T. et al. Digital twins for predictive oncology will be a paradigm shift for precision cancer care. *Nat. Med.* **27**, 2065–2066 (2021).
68. Coorey, G., Figtree, G. A., Fletcher, D. F. & Redfern, J. The health digital twin: advancing precision cardiovascular medicine. *Nat. Rev. Cardiol.* **18**, 803–804 (2021).
69. Masison, J. et al. A modular computational framework for medical digital twins. *Proc. Natl Acad. Sci. USA* **118**, e2024287118 (2021).

70. Fisher, C. K., Smith, A. M. & Walsh, J. R. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci. Rep.* **9**, 13622 (2019).
71. Walsh, J. R. et al. Generating digital twins with multiple sclerosis using probabilistic neural networks. Preprint at <https://arxiv.org/abs/2002.02779> (2020).
72. Swedish Digital Twin Consortium. <https://www.sdct.se/> (accessed 1 February 2022).
73. Potter, D. et al. Development of CancerLinQ, a health information learning platform from multiple electronic health record systems to support improved quality of care. *JCO Clin. Cancer Inform.* **4**, 929–937 (2020).
74. Parmar, P., Ryu, J., Pandya, S., Sedoc, J. & Agarwal, S. Health-focused conversational agents in person-centered care: a review of apps. *NPJ Digit. Med.* **5**, 21 (2022).
75. Dixon, R. F. et al. A virtual type 2 diabetes clinic using continuous glucose monitoring and endocrinology visits. *J. Diabetes Sci. Technol.* **14**, 908–911 (2020).
76. Claxton, S. et al. Identifying acute exacerbations of chronic obstructive pulmonary disease using patient-reported symptoms and cough feature analysis. *NPJ Digit. Med.* **4**, 107 (2021).
77. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
78. Patel, M. S., Volpp, K. G. & Asch, D. A. Nudge units to improve the delivery of health care. *N. Engl. J. Med.* **378**, 214–216 (2018).
79. Roller, S. et al. Recipes for building an open-domain Chatbot. In *Proc. 16th Conference of the European Chapter of the Association for Computational Linguistics* 300–325 (Association for Computational Linguistics, 2021).
80. Chen, J. H. & Asch, S. M. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N. Engl. J. Med.* **376**, 2507–2509 (2017).
81. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
82. Woodfield, R., Grant, I., UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group & Sudlow, C. L. M. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK biobank stroke outcomes group. *PLoS ONE* **10**, e0140533 (2015).
83. Szustakowski, J. et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* **53**, 942–948 (2021).
84. Halldórsson, B. V. et al. The sequences of 150,119 genomes in the UK Biobank. *Nature* **607**, 732–740 (2022).
85. Littlejohns, T. J. et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
86. Chen, Z. et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
87. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
88. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
89. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
90. All of Us Research Program Investigators. et al. The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
91. Mapes, B. M. et al. Diversity and inclusion for the All of Us research program: a scoping review. *PLoS ONE* **15**, e0234962 (2020).
92. Kaushal, A., Altman, R. & Langlotz, C. Geographic distribution of US cohorts used to train deep learning algorithms. *JAMA* **324**, 1212–1213 (2020).
93. Arges, K. et al. The Project Baseline Health Study: a step towards a broader mission to map human health. *NPJ Digit. Med.* **3**, 84 (2020).
94. McDonald, D. et al. American Gut: an open platform for citizen science microbiome research. *mSystems* **3**, e00031–18 (2018).
95. Johnson, A. E. W. et al. MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).
96. Johnson, A. E. W. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **6**, 317 (2019).
97. Deasy, J., Liò, P. & Ercole, A. Dynamic survival prediction in intensive care units from heterogeneous time series without the need for variable selection or curation. *Sci. Rep.* **10**, 22129 (2020).
98. Barbieri, S. et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. *Sci. Rep.* **10**, 1111 (2020).
99. Huang, S.-C., Pareek, A., Zamanian, R., Banerjee, I. & Lungren, M. P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci. Rep.* **10**, 22147 (2020).
100. Jabbour, S., Fouhey, D., Kazerooni, E., Wiens, J. & Sjoding, M. W. Combining chest X-rays and electronic health record data using machine learning to diagnose acute respiratory failure. *J. Am. Med. Inform. Assoc.* **29**, 1060–1068 (2022).
101. Golbus, J. R., Pescatore, N. A., Nallamothu, B. K., Shah, N. & Kheterpal, S. Wearable device signals and home blood pressure data across age, sex, race, ethnicity, and clinical phenotypes in the Michigan Predictive Activity & Clinical Trajectories in Health (MIPACT) study: a prospective, community-based observational study. *Lancet Digit. Health* **3**, e707–e715 (2021).
102. Addington, J. et al. North American Prodrome Longitudinal Study (NAPLS 2): overview and recruitment. *Schizophr. Res.* **142**, 77–82 (2012).
103. Perkins, D. O. et al. Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: preliminary results from the NAPLS project. *Schizophr. Bull.* **41**, 419–428 (2015).
104. Koutsouleris, N. et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* **78**, 195–209 (2021).
105. Baltrusaitis, T., Ahuja, C. & Morency, L.-P. Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443 (2019).
106. Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proc. 38th International Conference on Machine Learning* (eds. Meila, M. & Zhang, T.) vol. 139, 8748–8763 (PMLR, 18–24 July 2021).
107. Zhang, Y., Jiang, H., Miura, Y., Manning, C. D. & Langlotz, C. P. Contrastive learning of medical visual representations from paired images and text. Preprint at <https://arxiv.org/abs/2010.00747> (2020).
108. Zhou, H. -Y. et al. Generalized radiograph representation learning via cross-supervision between images and free-text radiology reports. *Nat. Mach. Intell.* **4**, 32–40 (2022).
109. Akbari, H. et al. VATT: transformers for multimodal self-supervised learning from raw video, audio and text. In *Advances in Neural Information Processing Systems* (eds. Ranzato, M. et al.) vol. 34, 24206–24221 (Curran Associates, Inc., 2021).
110. Bao, H. et al. VLMo: unified vision-language pre-training with mixture-of-modality-experts. Preprint at <https://arxiv.org/abs/2111.02358> (2022).
111. Dean, J. Introducing Pathways: a next-generation AI architecture <https://blog.google/technology/ai/introducing-pathways-next-generation-ai-architecture/> (10 November 2021).
112. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* (eds. Guyon, I. et al.) vol. 30 (Curran Associates, Inc., 2017).
113. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. In *International Conference on Learning Representations* (ICLR, 2021).
114. Li et al. Oscar: Object-semantics aligned pre-training for vision-language tasks. Preprint at <https://doi.org/10.48550/arXiv.2004.06165> (2020).
115. Baevski, A. et al. data2vec: a general framework for self-supervised learning in speech, vision and language. Preprint at <https://arxiv.org/abs/2202.03555> (2022).
116. Tamkin, A. et al. DABS: a Domain-Agnostic Benchmark for Self-Supervised Learning. In *35th Conf. Neural Information Processing Systems Datasets and Benchmarks Track* (2021).
117. Jaegle, A. et al. Perceiver: general perception with iterative attention. In *Proc. 38th International Conference on Machine Learning* (eds. Meila, M. & Zhang, T.) vol. 139, 4651–4664 (PMLR, 18–24 July 2021).
118. Jaegle, A. et al. Perceiver IO: a general architecture for structured inputs & outputs. In *International Conference on Learning Representations* (ICLR, 2022).
119. Hendricks, L. A., Mellor, J., Schneider, R., Alayrac, J.-B. & Nematzadeh, A. Decoupling the role of data, attention, and losses in multimodal transformers. *Trans. Assoc. Comput. Linguist.* **9**, 570–585 (2021).
120. Lu, K., Grover, A., Abbeel, P. & Mordatch, I. Pretrained transformers as universal computation engines. Preprint at <https://arxiv.org/abs/2103.05247> (2021).
121. Sandfort, V., Yan, K., Pickhardt, P. J. & Summers, R. M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci. Rep.* **9**, 16884 (2019).
122. Bai, X. et al. Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. *Nat. Mach. Intell.* **3**, 1081–1089 (2021).
123. Berisha, V. et al. Digital medicine and the curse of dimensionality. *NPJ Digit. Med.* **4**, 153 (2021).
124. Guu, K., Lee, K., Tung, Z., Pasupat, P. & Chang, M. Retrieval augmented language model pre-training. In *Proc. 37th International Conference on Machine Learning* (eds. Iii, H. D. & Singh, A.) vol. 119, 3929–3938 (PMLR, 13–18 July 2020).
125. Borgeaud, S. et al. Improving language models by retrieving from trillions of tokens. In *Proc. 39th International Conference on Machine Learning* (eds. Chaudhuri, K. et al.) vol. 162, 2206–2240 (PMLR, 17–23 July 2022).

126. Huang, S. -C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit. Med.* **3**, 136 (2020).
127. Muhammad, G. et al. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. *Inf. Fusion* **76**, 355–375 (2021).
128. Fiterau, M. et al. ShortFuse: Biomedical time series representations in the presence of structured information. In *Proc. 2nd Machine Learning for Healthcare Conference* (eds. Doshi-Velez, F. et al.) vol. 68, 59–74 (PMLR, 18–19 August 2017).
129. Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).
130. Rajpurkar, P. et al. CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest X-rays in patients with HIV. *NPJ Digit. Med.* **3**, 115 (2020).
131. Kihara, Y. et al. Policy-driven, multimodal deep learning for predicting visual fields from the optic disc and optical coherence tomography imaging. *Ophthalmology* <https://doi.org/10.1016/j.ophtha.2022.02.017> (2022).
132. Ramesh, A. et al. Zero-shot text-to-image generation. In *Proc. 38th International Conference on Machine Learning* (eds. Meila, M. & Zhang, T.) vol. 139, 8821–8831 (PMLR, 18–24 July 2021).
133. Nichol, A. Q. et al. GLIDE: towards photorealistic image generation and editing with text-guided diffusion models. In *Proc. 39th International Conference on Machine Learning* (eds. Chaudhuri, K. et al.) vol. 162, 16784–16804 (PMLR, 17–23 July 2022).
134. Reed, S. et al. A generalist agent. Preprint at <https://arxiv.org/abs/2205.06175> (2022).
135. Li, J. et al. Align before fuse: vision and language representation learning with momentum distillation. Preprint at <https://arxiv.org/abs/2107.07651> (2021).
136. Nagrani, A. et al. Attention bottlenecks for multimodal fusion. In *Advances in Neural Information Processing Systems* (eds. Ranzato, M. et al.) vol. 34, 14200–14213 (Curran Associates, Inc., 2021).
137. Hughes, J. W. et al. Deep learning evaluation of biomarkers from echocardiogram videos. *EBioMedicine* **73**, 103613 (2021).
138. Echle, A. et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer* **124**, 686–696 (2020).
139. Shilo, S., Rossman, H. & Segal, E. Axes of a revolution: challenges and promises of big data in healthcare. *Nat. Med.* **26**, 29–38 (2020).
140. Hripscak, G. et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud. Health Technol. Inform.* **216**, 574–578 (2015).
141. Rannikmäe, K. et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology* **95**, e697–e707 (2020).
142. Garg, R., Oh, E., Naidech, A., Kording, K. & Prabhakaran, S. Automating ischemic stroke subtype classification using machine learning and natural language processing. *J. Stroke Cerebrovasc. Dis.* **28**, 2045–2051 (2019).
143. Casey, B. J. et al. DSM-5 and RDoC: progress in psychiatry research? *Nat. Rev. Neurosci.* **14**, 810–814 (2013).
144. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
145. Zou, J. & Schiebinger, L. Ensuring that biomedical AI benefits diverse populations. *EBioMedicine* **67**, 103358 (2021).
146. Rocher, L., Hendrickx, J. M. & de Montjoye, Y. -A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat. Commun.* **10**, 3069 (2019).
147. Haneuse, S., Arterburn, D. & Daniels, M. J. Assessing missing data assumptions in EHR-based studies: a complex and underappreciated task. *JAMA Netw. Open* **4**, e210184–e210184 (2021).
148. van Smeden, M., Penning de Vries, B. B. L., Nab, L. & Groenwold, R. H. H. Approaches to addressing missing values, measurement error, and confounding in epidemiologic studies. *J. Clin. Epidemiol.* **131**, 89–100 (2021).
149. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
150. UK10K Consortium. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
151. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
152. Li, J. et al. Imputation of missing values for electronic health record laboratory data. *NPJ Digit. Med.* **4**, 147 (2021).
153. Tang, S. et al. Democratizing EHR analyses with FIDDLE: a flexible data-driven preprocessing pipeline for structured clinical data. *J. Am. Med. Inform. Assoc.* **27**, 1921–1934 (2020).
154. Che, Z. et al. Recurrent neural networks for multivariate time series with missing values. *Sci. Rep.* **8**, 6085 (2018).
155. Vokinger, K. N., Feuerriegel, S. & Kesselheim, A. S. Mitigating bias in machine learning for medicine. *Commun. Med.* **1**, 25 (2021).
156. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
157. Gichoya, J. W. et al. AI recognition of patient race in medical imaging: a modelling study. *Lancet Digit. Health* **4**, e406–e414 (2022).
158. Swanson, J. M. The UK Biobank and selection bias. *Lancet* **380**, 110 (2012).
159. Griffith, G. J. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11**, 5749 (2020).
160. Thompson, L. A. et al. The influence of selection bias on identifying an association between allergy medication use and SARS-CoV-2 infection. *EClinicalMedicine* **37**, 100936 (2021).
161. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
162. Keyes, K. M. & Westreich, D. UK Biobank, big data, and the consequences of non-representativeness. *Lancet* **393**, 1297 (2019).
163. Narayanan, A. & Shmatikov, V. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy* 111–125 (2008).
164. Gerke, S., Minssen, T. & Cohen, G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Artif. Intell. Health.* 11326, 213–227(2020).
165. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–311 (2020).
166. Rieke, N. et al. The future of digital health with federated learning. *NPJ Digit. Med.* **3**, 119 (2020).
167. Ziller, A. et al. Medical imaging deep learning with differential privacy. *Sci. Rep.* **11**, 13524 (2021).
168. Dayan, I. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat. Med.* **27**, 1735–1743 (2021).
169. Wood, A., Najarian, K. & Kahrobaei, D. Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Comput. Surv.* **53**, 1–35 (2020).
170. Warnat-Herresthal, S. et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature* **594**, 265–270 (2021).
171. Zhou, Z. et al. Edge intelligence: paving the last mile of artificial intelligence with edge computing. *Proc. IEEE* **107**, 1738–1762 (2019).
172. Intel. How edge computing is driving advancements in healthcare analytics; <https://www.intel.com/content/www/us/en/healthcare-it/edge-analytics.html> (11 March 2022).
173. Ballantyne, A. How should we think about clinical data ownership? *J. Med. Ethics* **46**, 289–294 (2020).
174. Liddell, K., Simon, D. A. & Lucassen, A. Patient data ownership: who owns your health? *J. Law Biosci.* **8**, lsab023 (2021).
175. Bierer, B. E., Crosas, M. & Pierce, H. H. Data authorship as an incentive to data sharing. *N. Engl. J. Med.* **376**, 1684–1687 (2017).
176. Scheibner, J. et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. *J. Med. Internet Res.* **23**, e25120 (2021).
177. Vamathevan, J. et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).

Acknowledgements

We thank A. Tamkin for invaluable feedback. NIH grant UL1TR002550 (to E.J.T.) supported this work.

Competing interests

Since completing this Review, J.N.A. became an employee of Rad AI. All the other authors declare no competing interests.

Additional information

Correspondence should be addressed to Pranav Rajpurkar or Eric J. Topol.

Peer review information *Nature Medicine* thanks Joseph Ledam, Leo Anthony Celi and Jenna Wiens for their contribution to the peer review of this work. Primary Handling Editor: Karen O'Leary, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2022