Review

Illuminating the dark spaces of healthcare with ambient intelligence

https://doi.org/10.1038/s41586-020-2669-y

Received: 16 February 2020

Accepted: 14 July 2020

Published online: 9 September 2020

Check for updates

Albert Haque[1], Arnold Milstein[2] & Li Fei-Fei[1,3]✉

Advances in machine learning and contactless sensors have given rise to ambient intelligence—physical spaces that are sensitive and responsive to the presence of humans. Here we review how this technology could improve our understanding of the metaphorically dark, unobserved spaces of healthcare. In hospital spaces, early applications could soon enable more efficient clinical workflows and improved patient safety in intensive care units and operating rooms. In daily living spaces, ambient intelligence could prolong the independence of older individuals and improve the management of individuals with a chronic disease by understanding everyday behaviour. Similar to other technologies, transformation into clinical applications at scale must overcome challenges such as rigorous clinical validation, appropriate data privacy and model transparency. Thoughtful use of this technology would enable us to understand the complex interplay between the physical environment and health-critical human behaviours.

Boosted by innovations in data science and artificial intelligence[1,2], decision-support systems are beginning to help clinicians to correct suboptimal and, in some cases, dangerous diagnostic and treatment decisions[3–5]. By contrast, the translation of better decisions into the physical actions performed by clinicians, patients and families remains largely unassisted[6]. Health-critical activities that occur in physical spaces, including hospitals and private homes, remain obscure. To gain the full dividends of medical advancements requires—in part—that affordable, human-centred approaches are continuously highlighted to assist clinicians in these metaphorically dark spaces.

Despite numerous improvement initiatives, such as surgical safety checklists[7], by the National Institutes of Health (NIH), Centres for Disease Control and Prevention (CDC), World Health Organization (WHO) and private organizations, as many as 400,000 people die every year in the United States owing to lapses and defects in clinical decision-making and physical actions[8]. Similar preventable suffering occurs in other countries, as well-motivated clinicians struggle with the rapidly growing complexity of modern healthcare[9,10]. To avoid overwhelming the cognitive capabilities of clinicians, advances in artificial intelligence hold the promise of assisting clinicians, not only with clinical decisions but also with the physical steps of clinical decisions[6].

Advances in machine learning and low-cost sensors can complement existing clinical decision-support systems by providing a computer-assisted understanding of the physical activities of healthcare. Passive, contactless sensors (Fig. 1) embedded in the environment can form an ambient intelligence that is aware of people's movements and adapt to their continuing health needs[11–14]. Similar to modern driver-assistance systems, this form of ambient intelligence can help clinicians and in-home caregivers to perfect the physical motions that comprise the final steps of modern healthcare. Already enabling better manufacturing, safer autonomous vehicles and smarter sports entertainment[15], clinical physical-action support can more reliably translate the rapid flow of biomedical discoveries into error-free healthcare delivery and worldwide human benefits.

This Review explores how ambient, contactless sensors, in addition to contact-based wearable devices, can illuminate two health-critical environments: hospitals and daily living spaces. With several illustrative clinical-use cases, we review recent algorithmic research and clinical validation studies, citing key patient outcomes and technical challenges. We conclude with a discussion of broader social and ethical considerations including privacy, fairness, transparency and ethics. Additional references can be found in Supplementary Note 1.

Hospital spaces

In 2018, approximately 7.4% of the US population required an overnight hospital stay[16]. In the same year, 17 million admission episodes were reported by the National Health Service (NHS) in the UK[17]. Yet, healthcare workers are often overworked, and hospitals understaffed and resource-limited[18,19]. We discuss a number of hospital spaces in which ambient intelligence may have an important role in improving the quality of healthcare delivery, the productivity of clinicians, and business operations (Fig. 2). These improvements could be of great assistance during healthcare crises, such as pandemics, during which time hospitals encounter a surge of patients[20].

Intensive care units

Intensive care units (ICUs) are specialized hospital departments in which patients with life-threatening illnesses or critical organ failures are treated. In the United States, ICUs cost the health system US$108 billion per year[21] and account for up to 13% of all hospital costs[22].

One promising use case of ambient intelligence in ICUs is the computer-assisted monitoring of patient mobilization. ICU-acquired weaknesses are a common neuromuscular impairment in critically ill

[1]Department of Computer Science, Stanford University, Stanford, CA, USA. [2]Clinical Excellence Research Center, Stanford University School of Medicine, Stanford, CA, USA. [3]Stanford Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA, USA. ✉e-mail: feifeili@stanford.edu

	Camera	Depth sensor	Thermal sensor	Radio sensor	Acoustic sensor
Sensory information	RGB, colour, video	Lidar	Infrared	Radar, Wi-Fi	Microphone
Function	Measures colour (visible light)	Measures distance to objects	Measures surface temperature	Estimates distance and velocity	Measures air pressure waves (sound)
Sampling rate	30 Hz (1,920 × 1,080)	30 Hz (1,280 × 720)	10 Hz (640 × 480)	800 Hz	44.1 kHz
Bit depth	24 bits	16 bits	16 bits	32 bits	16 bits
Uses	Object recognition, person detection	3D object detection, robotic navigation	Night vision, equipment safety	Motion detection, object detection	Speech recognition, event detection
Data visualization					

Fig. 1 | Contactless sensors for ambient intelligence. Brightly coloured pixels denote objects that are closer to the depth sensor. Black pixels denote sensor noise caused by reflective, metallic objects. The radio sensor shows a micro-Doppler signature of a moving object, for which the *x* axis denotes time (5 s) and the *y* axis denotes the Doppler frequency. The radio sensor image is reproduced from ref. [89]. The acoustic sensor displays an audio waveform of a person speaking, for which the *x* axis denotes time (5 s) and the *y* axis denotes the signal amplitude.

patients, potentially leading to a twofold increase in one-year mortality rate and 30% higher hospital costs[23]. Early patient mobilization could reduce the relative incidence of ICU-acquired weaknesses by 40%[24]. Currently, the standard mobility assessment is through direct, in-person observation, although its use is limited by cost impracticality, observer bias and human error[25]. Proper measurement requires a nuanced understanding of patient movements[26]. For example, localized wearable devices can detect pre-ambulation manoeuvres (for example, the transition from sitting to standing)[27], but are unable to detect external assistance or interactions with the physical space (for example, sitting on chair versus bed)[27]. Contactless, ambient sensors could provide the continuous and nuanced understanding needed to accurately measure patient mobility in ICUs.

In one pioneering study, researchers installed ambient sensors (Fig. 2a) in one ICU room (Fig. 2b) and collected 362 h of data from eight patients[28]. A machine-learning algorithm categorized in-bed, out-of-bed and walking activities with an accuracy of 87% when compared to retrospective review by three physicians. In a larger study at a different hospital (Fig. 2c), another research team installed depth sensors in eight ICU rooms[29]. They trained a convolutional neural network[1] on 379 videos to categorize mobility activities into four categories (Fig. 2d). When validated on an out-of-sample dataset of 184 videos, the algorithm demonstrated 87% sensitivity and 89% specificity. Although these preliminary results are promising, a more insightful evaluation could provide stratified results rather than aggregate performance on short, isolated video clips. For example, one study used cameras, microphones and accelerometers to monitor 22 patients in ICUs, with and without delirium, over 7 days[30]. The study found significantly fewer head motions of patients who were delirious compared with patients who were not. Future studies could leverage this technology to detect delirium sooner and provide researchers with a deeper understanding of how patient mobilization affects mortality, length of stay and patient recovery.

Another early application is the control of hospital infections. Worldwide, more than 100 million patients are affected by hospital-acquired (that is, nosocomial) infections each year[31], with up to 30% of patients in ICUs experiencing a nosocomial infection[32]. Proper compliance with hand hygiene protocols is one of the most effective methods of reducing the frequency of nosocomial infections[33]. However, measuring compliance remains challenging. Currently, hospitals rely on auditors to measure compliance, despite being expensive, non-continuous and biased[34]. Wearable devices, particularly radio-frequency identification (RFID) badges, are a potential solution. Unfortunately, RFID provides coarse location estimates (that is, within tens of centimetres[35]), making it unable to categorize fine-grained movements such as the WHO's five moments of hand hygiene[36]. Alternatively, ambient sensors could monitor handwashing activities with higher fidelity—differentiating true use of an alcohol-gel dispenser from a clinician walking near a dispenser. In a pioneering study, researchers installed depth sensors above wall-mounted dispensers across an entire hospital unit[37,38]. A deep-learning algorithm achieved an accuracy of 75% at measuring compliance for 351 handwashing events during one hour. During the same time period, an in-person observer was 63% accurate, while a proximity algorithm (for example, RFID) was only 18% accurate. In more nuanced studies, ambient intelligence detected the use of contact-precautions equipment[39] and physical contact with the patient[40]. A critical next step is to translate ambient observation into changes in clinical behaviour, with a goal of improving patient outcomes.

Operating rooms

Worldwide, more than 230 million surgical procedures are undertaken annually[41] with up to 14% of patients experiencing an adverse event[42]. This percentage could be reduced through quicker surgical feedback, such as more frequent coaching of technical skill, which could reduce the number of errors by 50%[43]. Currently, the skills of a surgeon are assessed by peers and supervisors[44], despite being time-consuming, infrequent and subjective. Wearable sensors can be attached to hands or instruments to estimate the surgeon's skills[45], but may inhibit hand dexterity or introduce sterilization complexity. Ambient cameras are an unobtrusive alternative[46]. One study trained a convolutional neural network[1] to track a needle driver in prostatectomy videos[47]. Using peer-evaluation as the reference standard, the algorithm categorized 12 surgeons into high- and low-skill groups with an accuracy of 92%. A different study used videos from ten cholecystectomy procedures to reconstruct the trajectories of instruments during surgery and linked them to technical ratings by expert surgeons[48]. Further studies, such as video-based surgical phase recognition[49], could potentially lead to improved surgical training. However, additional clinical validation is needed and appropriate feedback mechanisms must be tested.

In the operating room, ambient intelligence is not limited to endoscopic videos[50]. Another example is the surgical count—a process of counting used objects to prevent objects being accidentally retained inside the patient[51]. Currently, dedicated staff time and effort are required to visually and verbally count these objects. Owing to attention deficit and insufficient team communication[52], it is possible for the human-adjudicated count to incorrectly label an object as returned

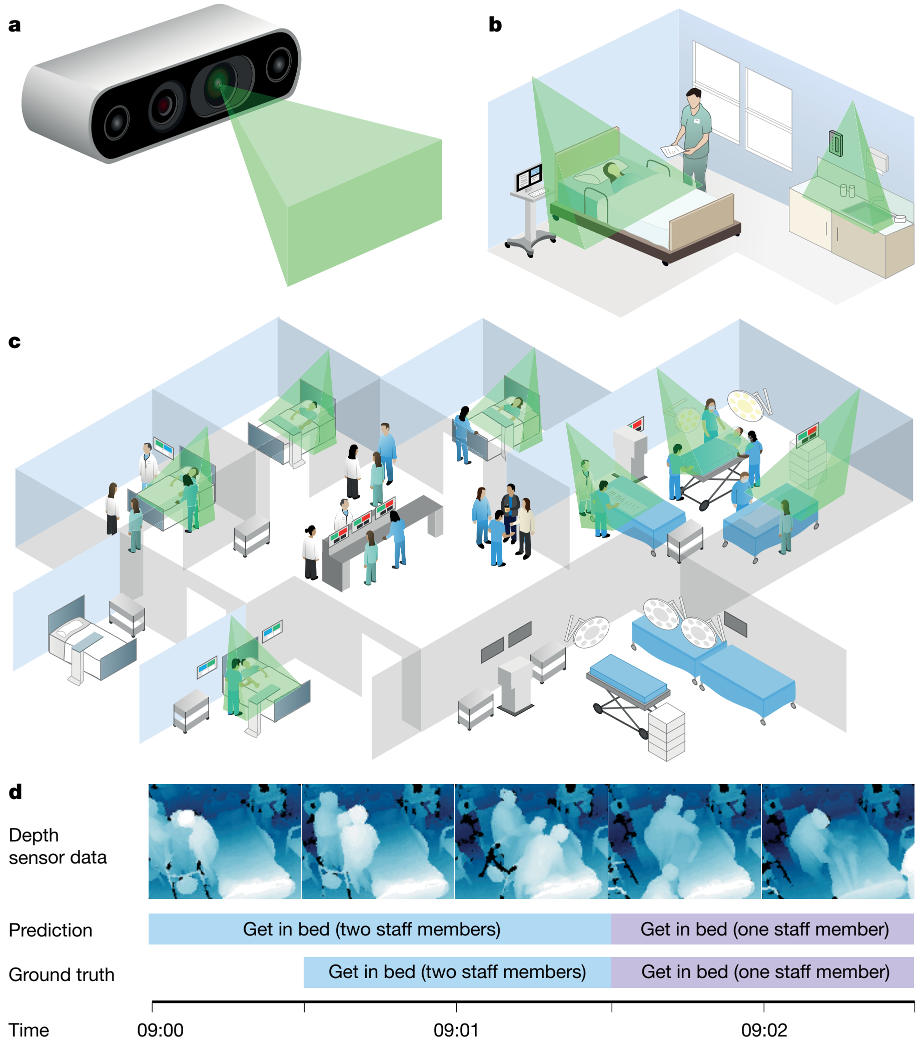


Fig. 2 | Ambient intelligence for hospitals. a, Commercial ambient sensor for which the coverage area is shown in green (that is, the field of view of visual sensors and range for acoustic and radio sensors). **b**, Sensors deployed inside a patient room can capture conversations and the physical motions of patients, clinicians and visitors. **c**, Sensors can be deployed throughout a hospital. **d**, Comparison of predictions and ground truth of activity from depth sensor data. Top, data from a depth sensor. Middle, the prediction of the algorithm of mobilization activity, duration and the number of staff who assist the patient. Bottom, human-annotated ground truth from a retrospective video review. **d**, Adapted from ref. [29].

when it is actually missing[51]. Automated counting systems, in particular, could assist surgical teams[53]. One study showed that barcode-equipped laparotomy sponges reduced the retained object rate from once every 16 days to once every 69 days[54]. Similar results were found with RFID and Raytec sponges[55]. However, owing to their size, barcodes and RFID cannot be applied to needles and instruments, which are responsible for up to 55% of counting discrepancies[51]—each discrepancy delaying the case by 13 min on average[51]. In addition to sponges, ambient cameras could count these smaller objects and potentially staff members[56]. In one operating room, researchers used ceiling-mounted cameras to track body parts of surgical team members with errors as low as five centimetres[57]. Ambient data collected throughout the room could create fine-grained logs of intraoperative activity[58]. Although these studies are promising as a proof of concept, further research needs to quantify the impact on patient outcomes, reimbursement and efficiency gains.

Other healthcare spaces

Clinicians spend up to 35% of their time on medical documentation tasks[59], taking valuable time away from patients. Currently, physicians perform documentation during or after each patient visit. Some providers use medical scribes to alleviate this burden, resulting in 0.17 more patients seen per hour and 0.21 more relative value units per patient (that is, insurer reimbursement)[60]. However, scribes are expensive to train and have high turnover[61]. Ambient microphones could perform a similar task to that of medical scribes[62]. Medical dictation software is an alternative, but is traditionally limited to the post-visit report[63]. In one study, researchers trained a deep-learning model on 14,000 h of outpatient audio from 90,000 conversations between patients and physicians[64]. The model demonstrated a word-level transcription accuracy of 80%, suggesting it may be better than the 76% accuracy of medical scribes[65]. In terms of clinical utility, one medical provider found that microphones attached to eyeglasses reduced time spent on documentation from 2 h to 15 min and doubled the time spent with patients[62].

From a management standpoint, ambient intelligence can improve the transition to activity-based costing[66]. Traditionally, insurance companies and hospital administrators estimated health outcomes per US dollar spent through a top-down approach of value-based accounting[67]. Time-driven activity-based costing is a bottom-up alternative and estimates the costs by individual resource time and cost (for example, the use of an ICU ventilator for 48 h)[68]. This can better inform process redesigns[66]—which, for one provider, led to 19% more patient visits with 17% fewer employees, without degradation of the patient outcomes[69]. Currently, in-person observations, staff interviews and electronic health records are used to map clinical activities to costs[68]. As described in this Review, ambient intelligence can automatically recognize clinical activities[70], count healthcare personnel[29] and estimate the duration of activities[29] (Fig. 2d). However, evidence of the clinical benefits of ambient intelligence is currently lacking, as the paradigm of activity-based costing is relatively new to hospital staff. As the technology develops, we hope that hospital administrators participate in the implementation and validation of ambient activity-based costing systems.

Daily living spaces

Humans spend a considerable portion of time at home. Around the world, the population is ageing[71]. Not only will this increase the amount of time spent at home, but it will also increase the importance of independent living, chronic disease management, physical rehabilitation and mental health of older individuals in daily living spaces.

Elderly living spaces and ageing

By 2050, the world's population aged 65 years or older will increase from 700 million to 1.5 billion[71]. Activities of daily living (ADLs), such as bathing, dressing and eating, are critical to the well-being and independence of this population. Impairment of one's ability to perform ADLs is associated with a twofold increase in falling risk[72] and up to a fivefold increase in one-year mortality rate[73]. Earlier detection of impairments could provide an opportunity to provide timely clinical care[11], potentially improving the ability to perform ADL by a factor of two[74]. Currently, ADLs are measured through self-reported questionnaires or manual grading by caregivers, despite the fact that these measurements are infrequent, biased and subjective[75]. Alternatively, wearable devices (such as accelerometers or electrocardiogram sensors) can track not only ADLs, but also heart rate, glucose level and respiration rate[76]. However, wearable devices are unable to discern whether a patient received ADL assistance—a key component of ADL evaluations[77]. Contactless, ambient sensors (Fig. 3a) could potentially identify these clinical nuances while detecting a greater range of activities[78].

In one of the first studies of its kind, researchers installed a depth and thermal sensor (Fig. 3b) inside the bedroom of an older individual and observed 1,690 activities during 1 month, including 231 instances of caregiver assistance[79] (Fig. 3c). A convolutional neural network[1] was 86% accurate at detecting assistance. In a different study, researchers collected ten days of video from six individuals in an elderly home and achieved similar results[80]. Although visual sensors are promising, they raise privacy concerns in some environments, such as bathrooms, which is where grooming, bathing and toileting activities occur, all of which are strongly indicative of cognitive function[81]. This led researchers to explore acoustic[82] and radar sensors[83]. One study used microphones to detect showering and toileting activities with accuracies of 93% and 91%, respectively[82]. However, a limitation of these studies is their evaluation in a small number of environments. Daily living spaces are highly variable, thus introducing generalization challenges. Additionally, privacy is of utmost importance. Development and verification of secure, privacy-safe systems is essential if this technology is to illuminate daily living spaces.

Another application for the independent living of older individuals is fall detection[84]. Approximately 29% of community-dwelling adults fall at least once a year[85]. Laying on the floor for more than one hour after a fall is correlated with a fivefold increase in 12-month mortality[86]. Furthermore, the fear of falling—associated with depression and lower quality of life[87]—can be reduced due to the perceived safety benefit of fall-detection systems[88]. For decades, researchers developed fall-detection systems with wearable devices and contactless ambient sensors[89]. A systematic review found that wearable devices detected falls with 96% accuracy while ambient sensors were 97% accurate[90]. In a different study, researchers installed Bluetooth (that is, radio) beacons in 271 homes[91]. Using signal strengths from each beacon, a machine-learning algorithm categorized the frailty of older individuals with an accuracy of 98%. In another study, researchers installed depth and radar sensors on the ceiling of 16 senior living apartments for 2 years[92]. Radar signals, transformed by a wavelet decomposition, detected 100% of falls with fewer than two false alarms per day[93]. Depth sensors produced one false alarm per month with a fall detection rate of 98%[94]. The ambient sensors were sufficiently fast (that is, low latency) to provide real-time email alerts to caregivers at 13 assisted-living communities[95]. Compared to a control group of 85 older individuals over 1 year, the real-time intervention significantly slowed the functional decline of 86 older individuals. When combined with wearable devices, one study found that the fall-detection accuracy of depth sensors increased from 90% to 98%[96], suggesting potential synergies between contactless and wearable sensors. As ambient intelligence begins to bridge the gap between observation and intervention, further studies are needed to explore regulatory approval processes, legal implications and ethical considerations.

Chronic disease management

With applications to physical rehabilitation and chronic diseases, gait analysis is an important tool for diagnostic testing and measuring treatment efficacy[97]. For example, frequent and accurate gait analysis could improve postoperative health for children with cerebral palsy[98] or enable earlier detection of Parkinson's disease by up to 4 years[99]. Traditionally limited to research laboratories with force plates and motion capture systems[100], gait analysis is being increasingly conducted with wearable devices[101]. One study used accelerometers to estimate the clinical-standard 6-min walking distance of 30 patients with chronic lung disease[102]. The study found an average absolute error rate of 6%. One limitation is that wearables must be physically attached to the body, making them inconvenient for patients[103]. Alternatively, contactless sensors could continuously measure gait with improved fidelity and create interactive, home-based rehabilitation programmes[104]. Several studies measured gait in natural settings with cameras[105], depth sensors[104], radar[106] and microphones[107]. One study used depth sensors to measure gait patterns of nine patients with Parkinson's disease[108]. Using a high-end motion capture system as the ground truth, the study found that depth sensors could track vertical knee motions to within four centimetres. Another study used depth sensors to create an exercise game for patients with cerebral palsy[109]. Over the course of 24 weeks, patients using the game improved their balance and gait by 18% according to the Tinetti test[110]. Although promising, these studies evaluated a single sensor modality. In laboratory experiments, gait detection improved by 3% to 7% when microphones were combined with wearable sensors[111]. When feasible, studies could investigate potential synergies of multiple sensing modalities (such as passive infrared motion sensors, contact sensors and wearable cameras).

Mental health

Mental illnesses, such as depression, anxiety and bipolar disorder, affect 43 million adults in the USA[112] and 165 million people in the European Union[113]. It is estimated that 56% of adults with mental illnesses do not seek treatment owing to barriers such as financial cost and provider availability[112]. Currently, self-reported questionnaires and clinical evaluations (for example, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)) are the standard tool for identifying symptoms of mental illness, despite being infrequent and biased[114]. Alternatively, ambient sensors could provide continuous and cost-effective symptom screening[115]. In one study, researchers collected audio, video and depth data from 69 individuals during 30-min, semi-structured clinical interviews[116]. Using the patient's verbal cues and upper body movement, a machine-learning algorithm detected 46 patients with schizophrenia with a positive predictive value of 95% and sensitivity of 84%. Similarly, in an emergency department, natural language analysis of clinical interviews with 61 adolescent individuals, of whom 31 were suicidal, yielded a model capable of categorizing patients who were suicidal with 90%

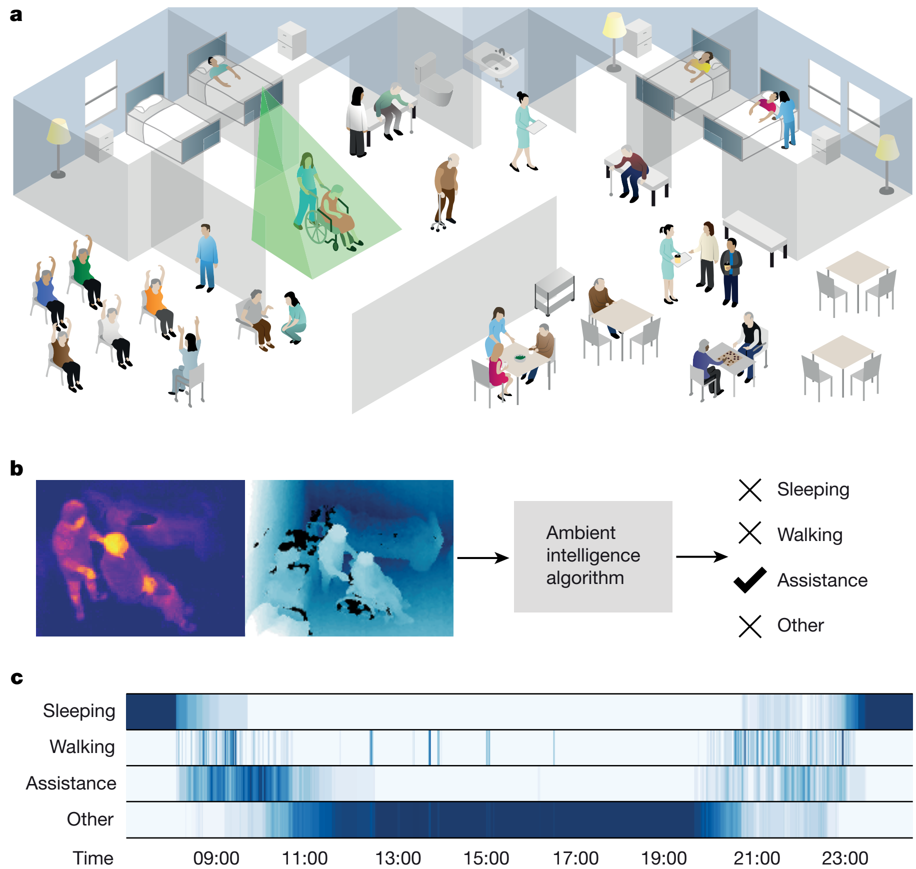


Fig. 3 | Ambient intelligence for daily living spaces. a, Elderly home equipped with one ambient sensor. The green frustum indicates the coverage area of the sensor (that is, the field of view for visual sensors and range for acoustic and radio sensors). **b**, Thermal and depth data from the sensor are processed by an ambient intelligence algorithm for activity categorization. **c**, Summary of a patient's activities for a single day. Darker blue sections indicate more frequent activity. **c**, Adapted from ref. [79].

accuracy[117]. Although impressive, further trials are needed to validate the effect on patient outcomes.

However, even after detection, treating mental health illnesses remains complex. Idiosyncratic therapist effects can cause up to 17% of the variance in outcomes[118], making it difficult to conduct psychotherapy research. Transcripts are the standard method for identifying features of good therapy[119], but are expensive to collect. Manual coding of a 20-min session can range from 85 to 120 min[120]. Ambient sensors could provide cheaper, higher-quality transcripts for psychotherapy research. Using text messages of treatment sessions, one study used a recurrent neural network[1] to detect instances of 24 therapist techniques from 14,899 patients[121]. The study identified several techniques correlated with improved Patient Health Questionnaire (PHQ-9) and General Anxiety Disorder (GAD-7) scores. A different study used microphones and a speech-recognition algorithm to transcribe and estimate therapists' empathy from 200 twenty-minute motivational interviewing sessions[120]. Using a committee of human assessors as the gold standard, the algorithm was 82% accurate. Although this is lower than the 90% accuracy of a single human assessor[120], ambient intelligence can more readily be applied to a larger number of patients. Using ambient intelligence, researchers can now conduct large-scale studies to reaffirm their understanding of psychotherapy frameworks. However, further research is needed to validate the generalization of these systems to a diverse population of therapists and patients.

Technical challenges and opportunities

Ambient intelligence can potentially illuminate the healthcare delivery process by observing recovery-related behaviours, reducing unintended clinician errors, assisting the ageing population and monitoring patients with chronic diseases. In Table 1, we highlight seven technical challenges and opportunities related to the recognition of human behaviour in complex scenes and learning from big data and rare events in clinical settings.

Behaviour recognition in complex scenes

Understanding complex human behaviours in healthcare spaces requires research that spans multiple areas of machine intelligence such as visual tracking, human pose estimation and human–object interaction models. Consider morning rounds in a hospital. Up to a dozen clinicians systematically review and visit each patient in a hospital unit. During this period, clinicians may occlude a sensor's view of the patient, potentially allowing health-critical activities to go undetected. If an object is moving before occlusion, tracking algorithms (Table 1) can estimate the position of the object while occluded[122]. For longer occlusions, matrix completion methods, such as image inpainting, can 'fill in' what is behind the occluding object[123]. Similar techniques can be used to denoise audio in spectrogram form[124]. If there are no occlusions, the next step is to locate people. During morning rounds, clinicians may hand each other objects or point across the room, introducing multiple layers of body parts from the perspective of the sensor. Human pose-estimation algorithms (Table 1) attempt to resolve this ambiguity by precisely locating body parts and assigning them to the correct individuals[125]. Building highly accurate human behaviour models is needed for ambient intelligence to succeed in complex clinical environments.

Ambient intelligence needs to understand how humans interact with objects and other people. One class of methods attempts to identify visually grounded relationships in images[126], commonly in the form of

Table 1 | Algorithmic challenges

Challenge	Sub-challenge	Technical approaches	ICUs		Operating rooms		Other		Elderly care		Chronic	Mental health	
			Patient mobility	Hand hygiene	Skills	Surgical count	Notes	Costing	ADLs	Falls	Gait analysis	Symptom screening	Therapy research
Behaviour recognition in complex scenes	Complex environments	Visual tracking, matrix completion	x	x	x	x	–	x	x	x	x	x	–
	Locating multiple humans	Pedestrian detection, human pose estimation	x	–	–	–	–	x	x	x	x	–	–
	Recognizing human behaviours	Scene graphs, activity recognition	x	x	–	x	–	x	x	x	–	–	x
Learning with big data and rare events	Big data	Distributed learning, optimizers	x	x	x	x	x	x	x	x	x	x	x
	Real-time detections	Two-stage models, model compression	–	x	x	x	–	–	–	x	–	–	–
	Rare events	Calibration, loss weighting	x	–	–	x	–	–	–	x	–	x	–
	Generalization to new environments	Transfer learning, few-shot learning	x	x	x	x	x	x	x	x	x	x	x

Rows denote algorithmic challenges. Columns denote clinical-use cases. Challenges applicable to specific clinical-use cases are marked by an 'x'. 'Skills' indicates the evaluation of surgical skills; 'notes' refers to medical documentation.

a scene graph (Table 1). A scene graph is a network of interconnected nodes, in which each node represents an object in the image and each connection represents their relationship[127]. Not only can scene graphs aid in the recognition of human behaviour, but they could also make ambient intelligence more transparent[128].

Learning from big data and rare events

Ambient sensors will produce petabytes of data from hospitals and homes[129]. This requires new machine-learning methods that are capable of modelling rare events and handling big data to be developed (Table 1). Large-scale activity-understanding models could require days to train unless large clusters of specialized hardware are used[130]. Cloud servers are a potential solution, but can be expensive as ambient intelligence may require considerable storage, computation and network bandwidth. Improved gradient-based optimizers[131] and neural network architectures[132] can potentially reduce training time. However, quickly training a model does not guarantee it will be fast during inference (that is, real-time detections) (Table 1). For example, video-based activity recognition models are slow, typically on the order of 1 to 10 frames per second[133]. Even optimized models capable of 100 frames per second[134] may have difficulties processing terabytes of data each day. Techniques such as model compression[135] and quantization[136] can reduce storage and computational requirements. Instead of processing audio or video at full spatial or temporal resolution, some methods quickly identify segments of interest, known as proposals[137]. These proposals are then provided to heavy-duty modules for highly accurate but computation-intensive activity recognition.

Although the volume of data produced by ambient sensors is large, some clinical events are rare and infrequent (Table 1). The detection of these long-tail events is necessary to understand health-critical behaviours. Consider the example of fall detection. The majority of ambient data contains normal activity, biasing the algorithm owing to label imbalance. More broadly, statistical bias can apply to any category of data, such as protected class attributes[138]. One solution is to statistically calibrate the algorithm, resulting in consistent error rates across specified attributes[139]. However, some healthcare environments may have a greater incidence of falls than in the original training set. This requires generalization (Table 1): the ability of an algorithm to operate on unseen distributions[140]. Instead of training a model designed for all distributions, one alternative is to take an existing model and fine-tune it on the new distribution[141]—also known as transfer learning[142]. Another solution, domain adaptation[143], attempts to minimize the gap between the training and testing distributions, often through better feature representations. For low-resource healthcare providers, few-shot learning—algorithms capable of learning from as few as one or two examples[144]—could be used.

Social and ethical considerations

Trustworthiness of ambient intelligence systems is critical to achieve the potential of this technology. Although there is an increasing body of literature on trustworthy artificial intelligence[145], we consider four separate dimensions of trustworthiness: privacy, fairness, transparency and research ethics. Developing the technology while addressing all four factors requires close collaborations between experts from medicine, computer science, law, ethics and public policy.

Privacy

Ambient sensors, by design, continuously observe the environment and can uncover new information about how physical human behaviours influence the delivery of healthcare. For example, sensors can measure vital signs from a distance[146]. While convenient, such knowledge could potentially be used to infer private medical conditions. As citizens worldwide are becoming more sensitive to mass data collection, there are growing concerns over confidentiality, sharing and retention of this information[147]. It is therefore essential to co-develop this technology with privacy and security in mind, not only in terms of the technology itself but also in terms of a continuous involvement of all stakeholders during the development[148].

A number of existing and emerging privacy-preserving techniques are presented in Fig. 4. One method is to de-identify data by removing the identities of the individuals. Another method is data minimization,

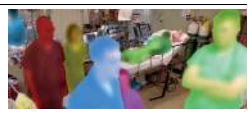
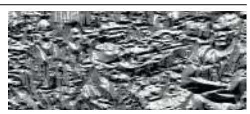
Method	Description	Computing hardware	Transformed result
Differential privacy	Adds noise to the data; minimally affects population-level analysis	Edge computer	
Face blurring	Detects and blurs human faces	Sensor, edge computer	
Dimensionality reduction	Reduces the input size by reducing the number of features	Sensor, edge computer	
Body masking	Replaces peoplewith faceless avatars	Edge computer	
Federated learning	Edge devices learn locally, then sends gradient updates to central server	Edge computer, centralized server	
Homomorphic encryption	Enables predictions to be made from encrypted data	Edge computer, centralized server	

Fig. 4 | Computational methods to protect privacy. There is a trade-off between the level of privacy protection provided by each method and the required computational resources. The methods used to generate the transformed images are described in detail elsewhere: differential privacy, ref. [166]; dimensionality reduction, ref. [167]; body masking, ref. [168]; federated learning, ref. [169]; homomorphic encryption, ref. [170]. The original image was produced by S. McCoy and has previously been published[171]. The appearance of US Department of Defence visual information does not imply or constitute endorsement by the US Department of Defence.

which minimizes data capture, transport and human bycatch. An ambient system could pause when a hospital room is unoccupied by a patient. However, even if data are de-identified, it may be possible to re-identify an individual[149]. Super-resolution techniques[150] can partially reverse the effects of face blurring and dimensionality reduction techniques, potentially enabling re-identification. This suggests that data should remain on-device to reduce the risk of unauthorized access and re-identification.

Legal and social complexities will inevitably arise. There are documented examples in which companies were required to provide data from ambient speakers and cameras to law enforcement[151]. Although these devices were located inside potential crime scenes, this raises the question at what point incidental findings outside the crime scene, such as inadvertent confessions, should be disclosed. Related to data sharing, some healthcare organizations have shared patient information with third parties such as data brokers[152]. To mitigate this, patients should proactively request healthcare providers to use privacy-preserving practices (Fig. 4). Additionally, clinicians and technologists must collaborate with critical stakeholders (for example, patients, family or caregivers), legal experts and policymakers to develop governance frameworks for ambient systems.

Fairness

Ambient intelligence will interact with large patient populations, potentially several orders of magnitude larger than the reach of current clinicians. This compels us to scrutinize the fairness of ambient systems. Fairness is a complex and multi-faceted topic, discussed by multiple research communities[138]. We highlight here two aspects of algorithmic fairness as examples: dataset bias and model performance.

Labelled datasets are the foundation of most machine-learning systems[1]. However, medical datasets have been biased, even before deep learning[153]. These biases can adversely affect clinical outcomes for certain populations[154]. If an individual is missing specific attributes, whether owing to data-collection constraints or societal factors, algorithms could misinterpret their entire record, resulting in higher levels of predictive error[155]. One method for identifying bias is to analyse model performance across different groups[156]. In one study, error rates varied across ethnic groups when predicting 30-day psychiatric readmission rates[157]. A more rigorous method could test for equal sensitivity and equal positive-predictive value. However, equal model performance may not produce equal clinical outcomes, as some populations may have inherent physiological differences. Nonetheless, progress is being made to mitigate bias, such as the PROBAST tool[158].

Transparency

Ambient intelligence can uncover insights about how healthcare delivery is influenced by human behaviour. These discoveries may surprise some researchers, in which case, clinicians and patients need to trust the findings before using them. Instead of opaque, black-box models, ambient intelligence systems should provide interpretable results that are predictive, descriptive and relevant[159]. This can aid in the challenging task of acquiring stakeholder buy-in, as technical illiteracy and model opacity can stagnate efforts to use ambient intelligence in healthcare[160]. Transparency is not limited to the algorithm. Dataset transparency—a detailed trace of how a dataset was designed, collected and annotated—would allow for specific precautions to be taken for future applications, such as training human annotators or revising the inclusion and exclusion criteria of a study. Formal guidelines on transparency, such as the TRIPOD statement[161], are actively being developed. Another tool is the use of model cards[162], which are short analyses that benchmark the algorithm across different populations and outline evaluation procedures.

Research ethics

Ethical research encompasses topics such as the protection of human participants, independent review and public beneficence. The Belmont Report, which prompted the regulation of research involving human participants, includes 'respect for persons' as a fundamental principle. In research, this manifests as informed consent from research participants. However, some regulations allow research to occur without consent if the research poses minimal risks to participants or if it is infeasible to obtain consent. For large-scale ambient intelligence studies, obtaining informed consent can be difficult, and it may in some cases be impossible due to automatic de-identification techniques (Fig. 4). In these cases, public engagement or deliberative democracy can be alternative solutions[163].

Relying solely on the integrity of principal investigators to conduct ethical research may introduce potential conflicts of interest. To mitigate this risk, academic research that involves human participants requires the approval from an Institutional Review Board. Public health surveillance, intended to prevent widespread disease and improve health, does not require independent review[164]. Depending on the application, ambient intelligence could be classified as either[165]. Researchers are urged to consult with experts from law and ethics to determine appropriate steps for protecting all human participants while maximizing public beneficence.

Summary

Centuries of medical practice led to a knowledge explosion, fuelling unprecedented advances in human health. Breakthroughs in artificial intelligence and low-cost, contactless sensors have given rise to an ambient intelligence that can potentially improve the physical execution of healthcare delivery. Preliminary results from hospitals and daily living spaces confirm the richness of information gained through ambient sensing. This extraordinary opportunity to illuminate the dark spaces of healthcare requires computer scientists, clinicians and medical researchers to work closely with experts from law, ethics and public policy to create trustworthy ambient intelligence systems for healthcare.

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
This paper reviews developments in deep learning and explains common neural network architectures such as convolutional and recurrent neural networks when applied to visual and natural language-processing tasks.
2. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
3. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med*. **25**, 24–29 (2019).
This perspective describes the use of computer vision, natural language processing, speech recognition and reinforcement learning for medical imaging tasks, electronic health record analysis, robotic-assisted surgery and genomic research.
4. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med*. **25**, 44–56 (2019).
This review outlines how artificial intelligence is used by clinicians, patients and health systems to interpret medical images, find workflow efficiencies and promote patient self-care.
5. Sutton, R. T. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit. Med*. **3**, 17 (2020).
6. Yeung, S., Downing, N. L., Fei-Fei, L. & Milstein, A. Bedside computer vision — moving artificial intelligence from driver assistance to patient safety. *N. Engl. J. Med*. **378**, 1271–1273 (2018).
7. Haynes, A. B. et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N. Engl. J. Med*. **360**, 491–499 (2009).
8. Makary, M. A. & Daniel, M. Medical error—the third leading cause of death in the US. *Br. Med. J*. **353**, i2139 (2016).
9. Tallentire, V. R., Smith, S. E., Skinner, J. & Cameron, H. S. Exploring error in team-based acute care scenarios: an observational study from the United Kingdom. *Acad. Med*. **87**, 792–798 (2012).
10. Yang, T. et al. Evaluation of medical malpractice litigations in China, 2002–2011. *J. Forensic Sci. Med*. **2**, 185–189 (2016).
11. Pol, M. C., ter Riet, G., van Hartingsveldt, M., Kröse, B. & Buurman, B. M. Effectiveness of sensor monitoring in a rehabilitation programme for older patients after hip fracture: a three-arm stepped wedge randomised trial. *Age Ageing* **48**, 650–657 (2019).
12. Fritz, R. L. & Dermody, G. A nurse-driven method for developing artificial intelligence in "smart" homes for aging-in-place. *Nurs. Outlook* **67**, 140–153 (2019).
13. Kaye, J. A. et al. F5-05-04: ecologically valid assessment of life activities: unobtrusive continuous monitoring with sensors. *Alzheimers Dement*. **12**, P374 (2016).
14. Acampora, G., Cook, D. J., Rashidi, P. & Vasilakos, A. V. A survey on ambient intelligence in health care. *Proc IEEE Inst. Electr. Electron. Eng*. **101**, 2470–2494 (2013).
15. Cook, D. J., Duncan, G., Sprint, G. & Fritz, R. Using smart city technology to make healthcare smarter. *Proc IEEE Inst. Electr. Electron. Eng*. **106**, 708–722 (2018).
16. Centers for Disease Control and Prevention. *National Health Interview Survey: Summary Health Statistics* https://www.cdc.gov/nchs/nhis/shs.htm (2018).
17. NHS Digital. *Hospital Admitted Patient Care and Adult Critical Care Activity 2018–19* https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2018-19 (NHS, 2019).
18. Patel, R. S., Bachu, R., Adikey, A., Malik, M. & Shah, M. Factors related to physician burnout and its consequences: a review. *Behav. Sci. (Basel)* **8**, 98 (2018).
19. Lyon, M. et al. Rural ED transfers due to lack of radiology services. *Am. J. Emerg. Med*. **33**, 1630–1634 (2015).
20. Adams, J. G. & Walls, R. M. Supporting the health care workforce during the COVID-19 global epidemic. *J. Am. Med. Assoc*. **323**, 1439–1440 (2020).
21. Halpern, N. A., Goldman, D. A., Tan, K. S. & Pastores, S. M. Trends in critical care beds and use among population groups and Medicare and Medicaid beneficiaries in the United States: 2000–2010. *Crit. Care Med*. **44**, 1490–1499 (2016).
22. Halpern, N. A. & Pastores, S. M. Critical care medicine in the United States 2000–2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit. Care Med*. **38**, 65–71 (2010).
23. Hermans, G. et al. Acute outcomes and 1-year mortality of intensive care unit-acquired weakness. A cohort study and propensity-matched analysis. *Am. J. Respir. Crit. Care Med*. **190**, 410–420 (2014).
24. Zhang, L. et al. Early mobilization of critically ill patients in the intensive care unit: a systematic review and meta-analysis. *PLoS ONE* **14**, e0223185 (2019).
25. Donchin, Y. et al. A look into the nature and causes of human errors in the intensive care unit. *Crit. Care Med*. **23**, 294–300 (1995).
26. Hodgson, C. L., Berney, S., Harrold, M., Saxena, M. & Bellomo, R. Clinical review: early patient mobilization in the ICU. *Crit. Care* **17**, 207 (2013).
27. Verceles, A. C. & Hager, E. R. Use of accelerometry to monitor physical activity in critically ill subjects: a systematic review. *Respir. Care* **60**, 1330–1336 (2015).
28. Ma, A. J. et al. Measuring patient mobility in the ICU using a novel noninvasive sensor. *Crit. Care Med*. **45**, 630–636 (2017).
29. Yeung, S. et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digit. Med*. **2**, 11 (2019).
This study used computer vision to simultaneously categorize patient mobilization activities in intensive care units and count the number of healthcare personnel involved in each activity.
30. Davoudi, A. et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci. Rep*. **9**, 8020 (2019).
This study used cameras and wearable sensors to track the physical movement of delirious and non-delirious patients in an intensive care unit.
31. WHO. *Report on the Burden of Endemic Health Care-associated Infection Worldwide* https://apps.who.int/iris/handle/10665/80135 (2011).
32. Vincent, J.-L. Nosocomial infections in adult intensive-care units. *Lancet* **361**, 2068–2077 (2003).
33. Gould, D. J., Moralejo, D., Drey, N., Chudleigh, J. H. & Taljaard, M. Interventions to improve hand hygiene compliance in patient care. *Cochrane Database Syst. Rev*. **9**, CD005186 (2017).
34. Srigley, J. A., Furness, C. D., Baker, G. R. & Gardam, M. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ Qual. Saf*. **23**, 974–980 (2014).
35. Shirehjini, A. A. N., Yassine, A. & Shirmohammadi, S. Equipment location in hospitals using RFID-based positioning system. *IEEE Trans. Inf. Technol. Biomed*. **16**, 1058–1069 (2012).
36. Sax, H. et al. 'My five moments for hand hygiene': a user-centred design approach to understand, train, monitor and report hand hygiene. *J. Hosp. Infect*. **67**, 9–21 (2007).
37. Haque, A. et al. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. In *Proc. 2nd Machine Learning for Healthcare Conference* 75–87 (PMLR, 2017).
This study evaluated the performance of depth sensors and covert auditors at measuring hand hygiene compliance in a hospital unit.
38. Singh, A. et al. Automatic detection of hand hygiene using computer vision technology. *J. Am. Med. Inform. Assoc*. https://doi.org/10.1093/jamia/ocaa115 (2020).
39. Chen, J., Cremer, J. F., Zarei, K., Segre, A. M. & Polgreen, P. M. Using computer vision and depth sensing to measure healthcare worker-patient contacts and personal protective equipment adherence within hospital rooms. *Open Forum Infect. Dis*. **3**, ofv200 (2016).
40. Awwad, S., Tarvade, S., Piccardi, M. & Gattas, D. J. The use of privacy-protected computer vision to measure the quality of healthcare worker hand hygiene. *Int. J. Qual. Health Care* **31**, 36–42 (2019).
41. Weiser, T. G. et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* **372**, 139–144 (2008).
42. Anderson, O., Davis, R., Hanna, G. B. & Vincent, C. A. Surgical adverse events: a systematic review. *Am. J. Surg*. **206**, 253–262 (2013).
43. Bonrath, E. M., Dedy, N. J., Gordon, L. E. & Grantcharov, T. P. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. *Ann. Surg*. **262**, 205–212 (2015).
44. Vaidya, A. et al. Current status of technical skills assessment tools in surgery: a systematic review. *J. Surg. Res*. **246**, 342–378 (2020).
45. Ghasemloonia, A. et al. Surgical skill assessment using motion quality and smoothness. *J. Surg. Educ*. **74**, 295–305 (2017).
46. Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B. & Rudzicz, F. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Netw. Open* **3**, e201664 (2020).
47. Law, H., Ghani, K. & Deng, J. Surgeon Technical skill assessment using computer vision based analysis. In *Proc. 2nd Machine Learning for Healthcare Conference* 88–99 (PMLR, 2017).
48. Jin, A. et al. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In *Proc. Winter Conference on Applications of Computer Vision* 691–699 (IEEE, 2018).
49. Twinanda, A. P. et al. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans. Med. Imaging* **36**, 86–97 (2017).
50. Hashimoto, D. A., Rosman, G., Rus, D. & Meireles, O. R. Artificial intelligence in surgery: promises and perils. *Ann. Surg*. **268**, 70–76 (2018).
51. Greenberg, C. C., Regenbogen, S. E., Lipsitz, S. R., Diaz-Flores, R. & Gawande, A. A. The frequency and significance of discrepancies in the surgical count. *Ann. Surg*. **248**, 337–341 (2008).
52. Agrawal, A. Counting matters: lessons from the root cause analysis of a retained surgical item. *Jt. Comm. J. Qual. Patient Saf*. **38**, 566–574 (2012).
53. Hempel, S. et al. Wrong-site surgery, retained surgical items, and surgical fires: a systematic review of surgical never events. *JAMA Surg*. **150**, 796–805 (2015).
54. Cima, R. R. et al. Using a data-matrix-coded sponge counting system across a surgical practice: impact after 18 months. *Jt. Comm. J. Qual. Patient Saf*. **37**, 51–58 (2011).
55. Rupp, C. C. et al. Effectiveness of a radiofrequency detection system as an adjunct to manual counting protocols for tracking surgical sponges: a prospective trial of 2,285 patients. *J. Am. Coll. Surg*. **215**, 524–533 (2012).
56. Kassahun, Y. et al. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *Int. J. Comput. Assist. Radiol. Surg*. **11**, 553–568 (2016).
57. Kadkhodamohammadi, A., Gangi, A., de Mathelin, M. & Padoy, N. A multi-view RGB-D approach for human pose estimation in operating rooms. In *Proc. Winter Conference on Applications of Computer Vision* 363–372 (IEEE, 2017).
58. Jung, J. J., Jüni, P., Lebovic, G. & Grantcharov, T. First-year analysis of the operating room black box study. *Ann. Surg*. **271**, 122–127 (2020).
59. Joukes, E., Abu-Hanna, A., Cornet, R. & de Keizer, N. F. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. *Appl. Clin. Inform*. **9**, 46–53 (2018).
60. Heaton, H. A., Castaneda-Guarderas, A., Trotter, E. R., Erwin, P. J. & Bellolio, M. F. Effect of scribes on patient throughput, revenue, and patient and provider satisfaction: a systematic review and meta-analysis. *Am. J. Emerg. Med*. **34**, 2018–2028 (2016).
61. Rich, N. The impact of working as a medical scribe. *Am. J. Emerg. Med*. **35**, 513 (2017).
62. Boulton, C. How Google Glass automates patient documentation for dignity health. *Wall Street Journal* (16 June 2014).

63. Blackley, S. V., Huynh, J., Wang, L., Korach, Z. & Zhou, L. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *J. Am. Med. Inform. Assoc*. **26**, 324–338 (2019).
64. Chiu, C.-C. et al. Speech recognition for medical conversations. In *Proc. 18th Annual Conference of the International Speech Communication Association* 2972–2976 (ISCA, 2018).
This paper developed a speech-recognition algorithm to transcribe anonymized conversations between patients and clinicians.
65. Pranaat, R. et al. Use of simulation based on an electronic health records environment to evaluate the structure and accuracy of notes generated by medical scribes: proof-of-concept study. *JMIR Med. Inform*. **5**, e30 (2017).
66. Kaplan, R. S. et al. Using time-driven activity-based costing to identify value improvement opportunities in healthcare. *J. Healthc. Manag*. **59**, 399–412 (2014).
67. Porter, M. E. Value-based health care delivery. *Ann. Surg*. **248**, 503–509 (2008).
68. Keel, G., Savage, C., Rafiq, M. & Mazzocato, P. Time-driven activity-based costing in health care: a systematic review of the literature. *Health Policy* **121**, 755–763 (2017).
69. French, K. E. et al. Measuring the value of process improvement initiatives in a preoperative assessment center using time-driven activity-based costing. *Healthcare* **1**, 136–142 (2013).
70. Sánchez, D., Tentori, M. & Favela, J. Activity recognition for the smart hospital. *IEEE Intelligent Systems* **23**, 50–57 (2008).
71. United Nations. *World Population Ageing 2019* https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2019_worldpopulationageing_report.pdf (2020).
72. Mamikonian-Zarpas, A. & Laganá, L. The relationship between older adults' risk for a future fall and difficulty performing activities of daily living. *J. Aging Gerontol*. **3**, 8–16 (2015).
73. Stineman, M. G. et al. All-cause 1-, 5-, and 10-year mortality in elderly people according to activities of daily living stage. *J. Am. Geriatr. Soc*. **60**, 485–492 (2012).
74. Phelan, E. A., Williams, B., Penninx, B. W. J. H., LoGerfo, J. P. & Leveille, S. G. Activities of daily living function and disability in older adults in a randomized trial of the health enhancement program. *J. Gerontol. A* **59**, M838–M843 (2004).
75. Carlsson, G., Haak, M., Nygren, C. & Iwarsson, S. Self-reported versus professionally assessed functional limitations in community-dwelling very old individuals. *Int. J. Rehabil. Res*. **35**, 299–304 (2012).
76. Wang, Z., Yang, Z. & Dong, T. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors* **17**, 341 (2017).
77. Katz, S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J. Am. Geriatr. Soc*. **31**, 721–727 (1983).
78. Uddin, M. Z., Khaksar, W. & Torresen, J. Ambient sensors for elderly care and independent living: a survey. *Sensors* **18**, 2027 (2018).
79. Luo, Z. et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. In *Proc. 3rd Machine Learning for Healthcare Conference* 1–18 (PMLR, 2018).
This study created spatial and temporal summaries of activities of daily living using a depth and thermal sensor inside the bedroom of an older resident.
80. Cheng, H., Liu, Z., Zhao, Y., Ye, G. & Sun, X. Real world activity summary for senior home monitoring. *Multimedia Tools Appl*. **70**, 177–197 (2014).
81. Lee, M.-T., Jang, Y. & Chang, W.-Y. How do impairments in cognitive functions affect activities of daily living functions in older adults? *PLoS ONE* **14**, e0218112 (2019).
82. Chen, J., Zhang, J., Kam, A. H. & Shue, L. An automatic acoustic bathroom monitoring system. In *Proc. International Symposium on Circuits and Systems* 1750–1753 (IEEE, 2005).
83. Shrestha, A. et al. Elderly care: activities of daily living classification with an S band radar. *J. Eng*. **2019**, 7601–7606 (2019).
84. Ganz, D. A. & Latham, N. K. Prevention of falls in community-dwelling older adults. *N. Engl. J. Med*. **382**, 734–743 (2020).
85. Bergen, G., Stevens, M. R. & Burns, E. R. Falls and fall injuries among adults aged ≥65 years — United States, 2014. *MMWR Morb. Mortal. Wkly. Rep*. **65**, 993–998 (2016).
86. Wild, D., Nayak, U. S. & Isaacs, B. How dangerous are falls in old people at home? *Br. Med. J. (Clin. Res. Ed.)* **282**, 266–268 (1981).
87. Scheffer, A. C., Schuurmans, M. J., van Dijk, N., van der Hooft, T. & de Rooij, S. E. Fear of falling: measurement strategy, prevalence, risk factors and consequences among older persons. *Age Ageing* **37**, 19–24 (2008).
88. Pol, M. et al. Older people's perspectives regarding the use of sensor monitoring in their home. *Gerontologist* **56**, 485–493 (2016).
89. Erol, B., Amin, M. G. & Boashash, B. Range-Doppler radar sensor fusion for fall detection. In *Proc. IEEE Radar Conference* 819–824 (IEEE, 2017).
90. Chaudhuri, S., Thompson, H. & Demiris, G. Fall detection devices and their use with older adults: a systematic review. *J. Geriatr. Phys. Ther*. **37**, 178–196 (2014).
91. Tegou, T. et al. A low-cost indoor activity monitoring system for detecting frailty in older adults. *Sensors* **19**, 452 (2019).
92. Rantz, M. et al. Automated in-home fall risk assessment and detection sensor system for elders. *Gerontologist* **55**, S78–S87 (2015).
93. Su, B. Y., Ho, K. C., Rantz, M. J. & Skubic, M. Doppler radar fall activity detection using the wavelet transform. *IEEE Trans. Biomed. Eng*. **62**, 865–875 (2015).
94. Stone, E. E. & Skubic, M. Fall detection in homes of older adults using the Microsoft Kinect. *IEEE J. Biomed. Health Inform*. **19**, 290–301 (2015).
95. Rantz, M. et al. Randomized trial of intelligent sensor system for early illness alerts in senior housing. *J. Am. Med. Dir. Assoc*. **18**, 860–870 (2017).
This randomized trial investigated the clinical efficacy of a real-time intervention system—triggered by abnormal gait patterns, as detected by ambient sensors—on the walking ability of older individuals at home.
96. Kwolek, B. & Kepski, M. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Comput. Methods Programs Biomed*. **117**, 489–501 (2014).
97. Wren, T. A. L., Gorton, G. E. III, Ounpuu, S. & Tucker, C. A. Efficacy of clinical gait analysis: a systematic review. *Gait Posture* **34**, 149–153 (2011).
98. Wren, T. A. et al. Outcomes of lower extremity orthopedic surgery in ambulatory children with cerebral palsy with and without gait analysis: results of a randomized controlled trial. *Gait Posture* **38**, 236–241 (2013).
99. Del Din, S. et al. Gait analysis with wearables predicts conversion to Parkinson disease. *Ann. Neurol*. **86**, 357–367 (2019).
100. Kidziński, Ł., Delp, S. & Schwartz, M. Automatic real-time gait event detection in children using deep neural networks. *PLoS ONE* **14**, e0211466 (2019).
101. Díaz, S., Stephenson, J. B. & Labrador, M. A. Use of wearable sensor technology in gait, balance, and range of motion analysis. *Appl. Sci*. **10**, 234 (2020).
102. Juen, J., Cheng, Q., Prieto-Centurion, V., Krishnan, J. A. & Schatz, B. Health monitors for chronic disease by gait analysis with mobile phones. *Telemed. J. E Health* **20**, 1035–1041 (2014).
103. Kononova, A. et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. *JMIR Mhealth Uhealth* **7**, e9832 (2019).
104. Da Gama, A., Fallavollita, P., Teichrieb, V., & Navab, N. Motor rehabilitation using Kinect: a systematic review. *Games Health J*. **4**, 123–135 (2015).
105. Cho, C.-W., Chao, W.-H., Lin, S.-H. & Chen, Y.-Y. A vision-based analysis system for gait recognition in patients with Parkinson's disease. *Expert Syst. Appl*. **36**, 7033–7039 (2009).
106. Seifert, A., Zoubir, A. M. & Amin, M. G. Detection of gait asymmetry using indoor Doppler radar. In *Proc. IEEE Radar Conference* 1–6 (IEEE, 2019).
107. Altaf, M. U. B., Butko, T., Juang, B. H. & Juang, B.-H. Acoustic gaits: gait analysis with footstep sounds. *IEEE Trans. Biomed. Eng*. **62**, 2001–2011 (2015).
108. Galna, B. et al. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait Posture* **39**, 1062–1068 (2014).
109. Jaume-i-Capó, A., Martínez-Bueso, P., Moyà-Alcover, B. & Varona, J. Interactive rehabilitation system for improvement of balance therapies in people with cerebral palsy. *IEEE Trans. Neural Syst. Rehabil. Eng*. **22**, 419–427 (2014).
110. Tinetti, M. E., Williams, T. F. & Mayewski, R. Fall risk index for elderly patients based on number of chronic disabilities. *Am. J. Med*. **80**, 429–434 (1986).
111. Wang, C. et al. Multimodal gait analysis based on wearable inertial and microphone sensors. In *Proc. IEEE SmartWorld* 1–8 (2017).
112. Mental Health America. *Mental Health in America - Adult Data 2018* https://www.mhanational.org/issues/mental-health-america-adult-data-2018 (2018).
113. Wittchen, H. U. et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol*. **21**, 655–679 (2011).
114. Snowden, L. R. Bias in mental health assessment and intervention: theory and evidence. *Am. J. Public Health* **93**, 239–243 (2003).
115. Shatte, A. B. R., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: a scoping review of methods and applications. *Psychol. Med*. **49**, 1426–1448 (2019).
116. Chakraborty, D. et al. Assessment and prediction of negative symptoms of schizophrenia from RGB+ D movement signals. In *Proc. 19th International Workshop on Multimedia Signal Processing* 1–6 (2017).
117. Pestian, J. P. et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. *Suicide Life Threat. Behav*. **46**, 154–159 (2016).
118. Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S. & Stiles, W. B. Therapist effects in outpatient psychotherapy: a three-level growth curve approach. *J. Couns. Psychol*. **54**, 32–39 (2007).
119. Miner, A. S. et al. Assessing the accuracy of automatic speech recognition for psychotherapy. *NPJ Digit. Med*. **3**, 82 (2020).
120. Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C. & Narayanan, S. S. "Rate my therapist": automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS ONE* **10**, e0143055 (2015).
121. Ewbank, M. P. et al. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry* **77**, 35–43 (2020).
122. Sadeghian, A., Alahi, A. & Savarese, S. Tracking the untrackable: learning to track multiple cues with long-term dependencies. In *Proc. Conference on Computer Vision and Pattern Recognition* 300–311 (IEEE, 2017).
123. Liu, G. et al. Image inpainting for irregular holes using partial convolutions. In *Proc. 15th European Conference on Computer Vision* 89–105 (Springer, 2018).
124. Marafioti, A., Perraudin, N., Holighaus, N. & Majdak, P. A context encoder for audio inpainting. *IEEE/ACM Trans. Audio Speech Lang. Process*. **27**, 2362–2372 (2019).
125. Chen, Y., Tian, Y. & He, M. Monocular human pose estimation: a survey of deep learning-based methods. *Comput. Vis. Image Underst*. **192**, 102897 (2020).
126. Krishna, R. et al. Visual genome: connecting language and vision using crowdsourced dense image annotations. *Int. J. Comput. Vis*. **123**, 32–73 (2017).
127. Johnson, J. et al. Image retrieval using scene graphs. In *Proc. Conference on Computer Vision and Pattern Recognition* 3668–3678 (IEEE, 2015).
128. Shi, J., Zhang, H. & Li, J. Explainable and explicit visual reasoning over scene graphs. In *Proc. Conference on Computer Vision and Pattern Recognition* 8368–8376 (IEEE/CVF, 2019).
129. Halamka, J. D. Early experiences with big data at an academic medical center. *Health Aff*. **33**, 1132–1138 (2014).
130. Verbraeken, J. et al. A survey on distributed machine learning. *ACM Comput. Surv*. **53**, 30 (2020).
131. You, Y. et al. Large batch optimization for deep learning: training BERT in 76 minutes. In *Proc. 8th International Conference on Learning Representations* 1–38 (2020).
132. Kitaev, N., Kaiser, Ł. & Levskaya, A. Reformer: the efficient transformer. In *Proc. 8th International Conference on Learning Representations* 1–12 (2020).
133. Heilbron, F., Niebles, J. & Ghanem, B. Fast temporal activity proposals for efficient detection of human actions in untrimmed videos. In *Proc. Conference on Computer Vision and Pattern Recognition* 1914–1923 (IEEE, 2016).

134. Zhu, Y., Lan, Z., Newsam, S. & Hauptmann, A. Hidden two-stream convolutional networks for action recognition. In *Proc. 14th Asian Conference on Computer Vision* 363–378 (Springer, 2019).
135. Han, S., Mao, H. & Dally, W. J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. In *Proc. 4th International Conference on Learning Representations* 1–14 (2016).
This paper introduced a method to compress neural network models and reduce their computational and storage requirements.
136. Micikeviciusd, P. et al. Mixed precision training. In *Proc. 6th International Conference on Learning Representations* 1–12 (2018).
137. Yu, G. & Yuan, J. Fast action proposals for human action detection and search. In *Proc. Conference on Computer Vision and Pattern Recognition* 1302–1311 (IEEE, 2015).
138. Zou, J. & Schiebinger, L. AI can be sexist and racist — it's time to make it fair. *Nature* **559**, 324–326 (2018).
139. Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J. & Weinberger, K. Q. On fairness and calibration. *Adv. Neural Inf. Process. Syst*. **30**, 5680–5689 (2017).
140. Neyshabur, B., Bhojanapalli, S., McAllester, D. & Srebro, N. Exploring generalization in deep learning. *Adv. Neural Inf. Process. Syst*. **30**, 5947–5956 (2017).
141. Howard, J. & Ruder, S. Universal language model fine-tuning for text classification. In *Proc. 56th Annual Meeting of the Association for Computational Linguistics* 328–339 (2018).
142. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng*. **22**, 1345–1359 (2010).
143. Patel, V. M., Gopalan, R., Li, R. & Chellappa, R. Visual domain adaptation: a survey of recent advances. *IEEE Signal Process. Mag*. **32**, 53–69 (2015).
144. Wang, Y., Kwok, J., Ni, L. M. & Yao, Q. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv*. **53**, 63 (2020).
145. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell*. **1**, 389–399 (2019).
146. Li, C., Lubecke, V. M., Boric-Lubecke, O. & Lin, J. A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring. *IEEE Trans. Microw. Theory Tech*. **61**, 2046–2060 (2013).
147. Rockhold, F., Nisen, P. & Freeman, A. Data sharing at a crossroads. *N. Engl. J. Med*. **375**, 1115–1117 (2016).
148. Wiens, J. et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med*. **25**, 1337–1340 (2019).
149. El Emam, K., Jonker, E., Arbuckle, L. & Malin, B. A systematic review of re-identification attacks on health data. *PLoS ONE* **6**, e28071 (2011).
150. Nasrollahi, K. & Moeslund, T. Super-resolution: a comprehensive survey. *Mach. Vis. Appl*. **25**, 1423–1468 (2014).
151. Brewster, T. How an amateur rap crew stole surveillance tech that tracks almost every American. *Forbes Magazine* (12 October 2018).
152. Cutler, J. E. How can patients make money off their medical data? *Bloomberg Law* (29 January 2019).
153. Cahan, E. M., Hernandez-Boussard, T., Thadaney-Israni, S. & Rubin, D. L. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digit. Med*. **2**, 78 (2019).
154. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med*. **169**, 866–872 (2018).
155. Char, D. S., Shah, N. H. & Magnus, D. Implementing machine learning in health care — addressing ethical challenges. *N. Engl. J. Med*. **378**, 981–983 (2018).
156. Buolamwini, J. & Gebru, T. Gender shades: intersectional accuracy disparities in commercial gender classification. In *Proc. 1st Conference on Fairness, Accountability and Transparency* 77–91 (2018).
157. Chen, I. Y., Szolovits, P. & Ghassemi, M. Can AI help reduce disparities in general medical and mental health care? *AMA J. Ethics* **21**, E167–E179 (2019).
158. Wolff, R. F. et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med*. **170**, 51–58 (2019).
159. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl Acad. Sci. USA* **116**, 22071–22080 (2019).
This article proposed a framework for evaluating model interpretability through predictive accuracy, descriptive accuracy and relevancy.
160. He, J. et al. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med*. **25**, 30–36 (2019).
161. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann. Intern. Med*. **162**, 55–63 (2015).
162. Mitchell, M. et al. Model cards for model reporting. In *Proc. 2nd Conference on Fairness, Accountability, and Transparency* 220–229 (2019).
163. Thomas, R. et al. Deliberative democracy and cancer screening consent: a randomised control trial of the effect of a community jury on men's knowledge about and intentions to participate in PSA screening. *BMJ Open* **4**, e005691 (2014).
164. Otto, J. L., Holodniy, M. & DeFraites, R. F. Public health practice is not research. *Am. J. Public Health* **104**, 596–602 (2014).
165. Gerke, S., Yeung, S. & Cohen, I. G. Ethical and legal aspects of ambient intelligence in hospitals. *J. Am. Med. Assoc*. **323**, 601–602 (2020).
166. Kim, J. W., Jang, B. & Yoo, H. Privacy-preserving aggregation of personal health data streams. *PLoS ONE* **13**, e0207639 (2018).
167. van der Maaten, L., Postma, E. & van den Herik, J. Dimensionality reduction: a comparative. *J. Mach. Learn. Res*. **10**, 13 (2009).
168. Kocabas, M., Athanasiou, N. & Black, M. J. VIBE: video inference for human body pose and shape estimation. In *Proc. Conference on Computer Vision and Pattern Recognition* 5253–5263 (IEEE/CVF, 2020).
169. McMahan, H. B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Proc. 20th International Conference on Artificial Intelligence and Statistics* 1273–1282 (PMLR, 2017).
This paper proposed federated learning, a method for training a shared model while the data is distributed across multiple client devices.
170. Gentry, C. Fully homomorphic encryption using ideal lattices. In *Proc 41st Symposium on Theory of Computing* 169–178 (ACM, 2009).
This paper proposed the first fully homomorphic encryption scheme that supports addition and multiplication on encrypted data.
171. McCoy, S. T. *Aboard USNS Comfort* (US Navy, 2003).

Acknowledgements We thank A. Kaushal, D. C. Magnus, G. Burke, K. Schulman and M. Hutson for providing comments on this paper. We also thank our clinical collaborators over the years, including A. S. Miner, A. Singh, B. Campbell, D. F. Amanatullah, F. R. Salipur, H. Rubin, J. Jopling, K. Deru, N. L. Downing, R. Nazerali, T. Platchek and W. Beninati, and our technical collaborators over the years, including A. Alahi, A. Rege, B. Liu, B. Peng, D. Zhao, E. Chou, E. Adeli, G. M. Bianconi, G. Pusiol, H. Cai, J. Beal, J.-T. Hsieh, M. Guo, R. Mehra, S. Mehra, S. Yeung and Z. Luo. A.H.'s graduate work was partially supported by the US Office of Naval Research (grant N00014-16-1-2127) and the Stanford Institute for Human-Centered Artificial Intelligence.

Author contributions A.H., A.M. and L.F-F. conceptualized the paper and its structure. A.H. and L.F.-F. wrote the paper. A.H. created the figures. A.M. provided substantial additions and edits. All authors contributed to multiple parts of the paper, as well as the final style and overall content.

Competing interests A.M. has financial interests in Prealize Health. L.F.-F. and A.M. have financial interests in Dawnlight Technologies. A.H. declares no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2669-y.
Correspondence and requests for materials should be addressed to L.F.-F.
Peer review information *Nature* thanks Andrew Beam, Eric Topol and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.