REVIEW ARTICLE   OPEN

Deep learning-enabled medical computer vision

Andre Esteva [1 ✉], Katherine Chou[2,5], Serena Yeung[3,5], Nikhil Naik [1,5], Ali Madani[1,5], Ali Mottaghi[3,5], Yun Liu [2], Eric Topol[4], Jeff Dean[2] and Richard Socher[1]

A decade of unprecedented progress in artificial intelligence (AI) has demonstrated the potential for many fields—including medicine—to benefit from the insights that AI techniques can extract from data. Here we survey recent progress in the development of modern computer vision techniques—powered by deep learning—for medical applications, focusing on medical imaging, medical video, and clinical deployment. We start by briefly summarizing a decade of progress in convolutional neural networks, including the vision tasks they enable, in the context of healthcare. Next, we discuss several example medical imaging applications that stand to benefit—including cardiology, pathology, dermatology, ophthalmology–and propose new avenues for continued work. We then expand into general medical video, highlighting ways in which clinical workflows can integrate computer vision to enhance care. Finally, we discuss the challenges and hurdles required for real-world clinical deployment of these technologies.

INTRODUCTION

Computer vision (CV) has a rich history spanning decades[1] of efforts to enable computers to perceive visual stimuli meaningfully. Machine perception spans a range of levels, from low-level tasks such as identifying edges, to high-level tasks such as understanding complete scenes. Advances in the last decade have largely been due to three factors: (1) the maturation of deep learning (DL)—a type of machine learning that enables end-to-end learning of very complex functions from raw data[2] (2) strides in localized compute power via GPUs[3], and (3) the open-sourcing of large labeled datasets with which to train these algorithms[4]. The combination of these three elements has enabled individual researchers the resource access needed to advance the field. As the research community grew exponentially, so did progress.

The growth of modern CV has overlapped with the generation of large amounts of digital data in a number of scientific fields. Recent medical advances have been prolific[5,6], owing largely to DL's remarkable ability to learn many tasks from most data sources. Using large datasets, CV models can acquire many pattern-recognition abilities—from physician-level diagnostics[7] to medical scene perception[8]. See Fig. 1.

Here we survey the intersection of CV and medicine, focusing on research in medical imaging, medical video, and real clinical deployment. We discuss key algorithmic capabilities which unlocked these opportunities, and dive into the myriad of accomplishments from recent years. The clinical tasks suitable for CV span many categories, such as screening, diagnosis, detecting conditions, predicting future outcomes, segmenting pathologies from organs to cells, monitoring disease, and clinical research. Throughout, we consider the future growth of this technology and its implications for medicine and healthcare.

COMPUTER VISION

Object classification, localization, and detection, respectively refer to identifying the type of an object in an image, the location of objects present, and both type and location simultaneously.

The ImageNet Large-Scale Visual Recognition Challenge[9] (ILSVRC) was a spearhead to progress in these tasks over the last decade. It created a large community of DL researchers competing and collaborating together to improve techniques on various CV tasks. The first contemporary, GPU-powered DL approach, in 2012[10], yielded an inflection point in the growth of this community, heralding an era of significant year-over-year improvements[11–14] through the competition's final year in 2017. Notably, classification accuracy achieved *human-level* performance during this period. Within medicine, fine-grained versions of these methods[15] have successfully been applied to the classification and detection of many diseases (Fig. 2). Given sufficient data, the accuracy often matches or surpasses the level of expert physicians[7,16]. Similarly, the *segmentation* of objects has substantially improved[17,18], particularly in challenging scenarios such as the biomedical segmentation of multiple types of overlapping cells in microscopy. The key DL technique leveraged in these tasks is the convolutional neural network[19] (CNN)—a type of DL algorithm which hardcodes translational invariance, a key feature of image data. Many other CV tasks have benefited from this progress, including image registration (identifying corresponding points across similar images), image retrieval (finding similar images), and image reconstruction and enhancement. The specific challenges of working with medical data require the utilization of many types of AI models.

These techniques largely rely on supervised learning, which leverages datasets that contain both data points (e.g. images) and data labels (e.g. object classes). Given the sparsity and access difficulties of medical data, transfer learning—in which an algorithm is first trained on a large and unrelated corpus (e.g. ImageNet[4]), then fine-tuned on a dataset of interest (e.g. medical)—has been critical for progress. To reduce the costs associated with collecting and labeling data, techniques to generate synthetic data, such as data augmentation[20] and generative adversarial networks (GANs)[21] are being developed. Researchers have even shown that crowd-sourcing image annotations can yield effective medical algorithms[22,23]. Recently, self-supervised

[1]Salesforce AI Research, San Francisco, CA, USA. [2]Google Research, Mountain View, CA, USA. [3]Stanford University, Stanford, CA, USA. [4]Scripps Research Translational Institute, La Jolla, CA, USA. [5]These authors contributed equally: Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi. ✉email: andre.esteva@gmail.com

learning[24]—in which *implicit* labels are extracted from data points and used to train algorithms (e.g predicting the spatial arrangement of tiles generated from splitting an image into pieces)—have pushed the field towards fully unsupervised learning, which lacks the need for labels. Applying these techniques in medicine will reduce the barrier to development and deployment.

Medical data access is central to this field, and key ethical and legal questions must be addressed. Do patients own their de-identified data? What if methods to re-identify data improve over time? Should the community open-source large quantities of data? To date, academia and industry have largely relied on small, open-source datasets, and data collected through commercial products. Dynamics around data sharing and country-specific availability will impact deployment opportunities. The field of *federated* learning[25]—in which centralized algorithms can be trained on distributed data that never leaves protected enclosures—may enable a workaround in stricter jurisdictions.
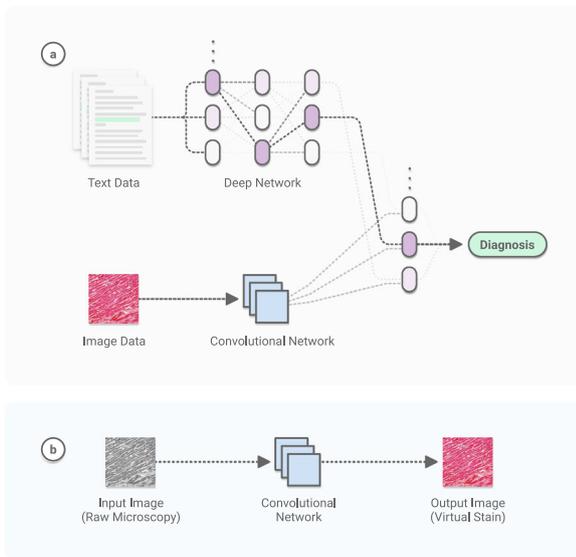
These advances have spurred growth in other domains of CV, such as multimodal learning, which combines vision with other modalities such as language (Fig. 1a)[26], time-series data, and genomic data[5]. These methods can combine with 3D vision[27,28] to turn depth-cameras into privacy-preserving sensors[29], making deployment easier for patient settings such as the intensive care unit[8]. The range of tasks is even broader in video. Applications like activity recognition[30] and live scene understanding[31] are useful in detecting and responding to important or adverse clinical events[32].

MEDICAL IMAGING

In recent years the number of publications applying computer vision techniques to static medical imagery has grown from hundreds to thousands[33]. A few areas have received substantial attention—radiology, pathology, ophthalmology, and dermatology—owing to the visual pattern-recognition nature of diagnostic tasks in these specialities, and the growing availability of highly structured images.

The unique characteristics of medical imagery pose a number of challenges to DL-based computer vision. For one, images can be massive. Digitizing histopathology slides produces gigapixel images of around 100,000 ×100,000 pixels, whereas typical CNN image inputs are around 200 ×200 pixels. Further, different chemical preparations will render different slides for the same piece of tissue, and different digitization devices or settings may produce different images for the same slide. Radiology modalities such as CT and MRI render equally massive 3D images, forcing standard CNNs to either work with a set of 2D slices, or adjust their internal structure to process in 3D. Similarly, ultrasound renders a time-series of noisy 2D slices of a 3D context–slices which are spatially correlated but not aligned. DL has started to account for the unique challenges of medical data. For instance, multiple-instance-learning (MIL)[34] enables learning from datasets containing massive images and few labels (e.g. histopathology). 3D convolutions in CNNs are enabling better learning from 3D volumes (e.g MRI and CT)[35]. Spatio-temporal models[36] and image registration enable working with time-series images (e.g. ultrasound).

Dozens of companies have obtained US FDA and European CE approval for medical imaging AI[37], and commercial markets have begun to form as sustainable business models are created. For instance, regions of high-throughput healthcare, such as India and Thailand, have welcomed the deployment of technologies such as diabetic retinopathy screening systems[38]. This rapid growth has now reached the point of directly impacting patient outcomes—the US CMS recently approved reimbursement for a radiology stroke triage use-case which reduces the time it takes for patients to receive treatment[39].

Fig. 1 Example medical computer vision tasks. a Multimodal discriminative model. Deep learning architectures can be constructed to jointly learn from both image data, typically with convolutional networks, and non-image data, typically with general deep networks. Learned annotations can include disease diagnostics, prognostics, clinical predictions, and combinations thereof. **b** Generative model. Convolutional neural networks can be trained to generate images. Tasks include image-to-image regression (shown), super-resolution image enhancement, novel image generation, and others.
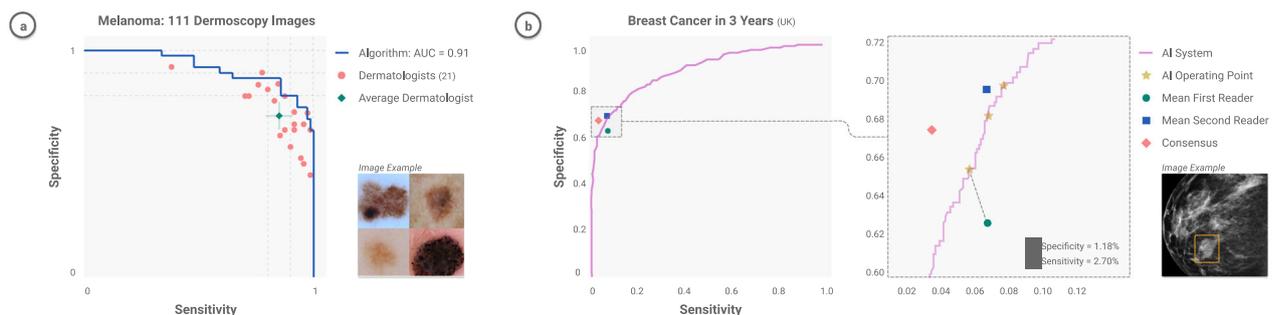

Fig. 2 Physician-level diagnostic performance. CNNs—trained to classify disease states—have been extensively tested across diseases, and benchmarked against physicians. Their performance is typically on par with experts when both are tested on the same image classification task. **a** Dermatology[7] and **b** Radiology[156]. Examples reprinted with permission and adapted for style.

CV in medical modalities with non-standardized data collection requires the integration of CV into existing physical systems. For instance, in otolaryngology, CNNs can be used to help primary care physicians manage patients' ears, nose, and throat[40], through mountable devices attached to smartphones[41]. Hematology and serology can benefit from microscope-integrated AIs[42] that diagnose common conditions[43] or count blood cells of various types[44]—repetitive tasks that are easy to augment with CNNs. AI in gastroenterology has demonstrated stunning capabilities. Video-based CNNs can be integrated into endoscopic procedures[45] for scope guidance, lesion detection, and lesion diagnosis. Applications include esophageal cancer screening[46], detecting gastric cancer[47,48], detecting stomach infections such as *H. Pylori*[49], and even finding hookworms[50]. Scientists have taken this field one step further by building entire medical AI devices designed for monitoring, such as at-home smart toilets outfitted with diagnostic CNNs on cameras[51]. Beyond the analysis of disease states, CV can serve the future of human health and welfare through applications such as screening human embryos for implantation[52].

Computer vision in radiology is so pronounced that it has quickly burgeoned into its own field of research, growing a corpus of work[53–55] that extends into all modalities, with a focus on X-rays, CT, and MRI. Chest X-ray analysis—a key clinical focus area[33]—has been an exemplar. The field has collected nearly 1 million annotated, open-source images[56–58]—the closest ImageNet[9] equivalent to date in medical CV. Analysis of brain imagery[59] (particularly for time-critical use-cases like stroke), and abdominal imagery[60] have similarly received substantial attention. Disease classification, nodule detection[61], and region segmentation (e.g. ventricular[62]) models have been developed for most conditions for which data can be collected. This has enabled the field to respond rapidly in times of crisis—for instance, developing and deploying COVID-19 detection models[63]. The field continues to expand with work in image translation (e.g. converting noisy ultrasound images into MRI), image reconstruction and enhancement (e.g. converting low-dosage, low-resolution CT images into high-resolution images[64]), automated report generation, and temporal tracking (e.g. image registration to track tumor growth over time). In the sections below, we explore vision-based applications in other specialties.

CARDIOLOGY
Cardiac imaging is increasingly used in a wide array of clinical diagnoses and workflows. Key clinical applications for deep learning include diagnosis and screening. The most common imaging modality in cardiovascular medicine is the cardiac ultrasound, or echocardiogram. As a cost-effective, radiation-free technique, echocardiography is uniquely suited for DL due to straightforward data acquisition and interpretation—it is routinely used in most acute inpatient facilities, outpatient centers, and emergency rooms[65]. Further, 3D imaging techniques such as CT and MRI are used for the understanding of cardiac anatomy and to better characterize supply-demand mismatch. CT segmentation algorithms have even been FDA—cleared for coronary artery visualization[66].

There are many example applications. DL can be trained on a large database of echocardiographic studies and surpass the performance of board-certified echocardiographers in view classification[67]. Computational DL pipelines can assess hypertrophic cardiomyopathy, cardiac amyloid, and pulmonary arterial hypertension[68]. EchoNet[69]—a deep learning model that can recognize cardiac structures, estimate function, and predict systemic phenotypes that are not readily identifiable to human interpretation—has recently furthered the field.

To account for challenges around data access,[70] data-efficient echocardiogram algorithms[70] have been developed, such as semi-supervised GANs that are effective at downstream tasks (e.g predicting left ventricular hypertrophy). To account for the fact that

most studies utilize privately held medical imaging datasets, 10,000 annotated echocardiogram videos were recently open-sourced[36]. Alongside this release, a video-based model, EchoNet-Dynamic[36], was developed. It can estimate ejection fraction and assess cardiomyopathy, alongside a comprehensive evaluation criterion based on results from an external dataset and human experts.

PATHOLOGY
Pathologists play a key role in cancer detection and treatment. Pathological analysis—based on visual inspection of tissue samples under microscope—is inherently subjective in nature. Differences in visual perception and clinical training can lead to inconsistencies in diagnostic and prognostic opinions[71–73]. Here, DL can support critical medical tasks, including diagnostics, prognostication of outcomes and treatment response, pathology segmentation, disease monitoring, and so forth.

Recent years have seen the adoption of sub-micron-level resolution tissue scanners that capture gigapixel whole-slide images (WSI)[74]. This development, coupled with advances in CV has led to research and commercialization activity in AI-driven digital histopathology[75]. This field has the potential to (i) overcome limitations of human visual perception and cognition by improving the efficiency and accuracy of routine tasks, (ii) develop new signatures of disease and therapy from morphological structures invisible to the human eye, and (iii) combine pathology with radiological, genomic, and proteomic measurements to improve diagnosis and prognosis[76].

One thread of research has focused on automating the routine, time-consuming task of localization and quantification of morphological features. Examples include the detection and classification of cells, nuclei, and mitoses[77–79], and the localization and segmentation of histological primitives such as nuclei, glands, ducts, and tumors[80–83]. These methods typically require expensive manual annotation of tissue components by pathologists as training data.

Another research avenue focuses on direct diagnostics[84–86] and prognostics[87,88] from WSI or tissue microarrays (TMA) for a variety of cancers—breast, prostate, lung cancer, etc. Studies have even shown that morphological features captured by a hematoxylin and eosin (H&E) stain are predictive of molecular biomarkers utilized in theragnosis[85,89]. While histopathology slides digitize into massive, data-rich gigapixel images, region-level annotations are sparse and expensive. To help overcome this challenge, the field has developed DL algorithms based on multiple-instance learning[90] that utilize slide-level "weak" annotations and exploit the sheer size of these images for improved performance.

The data abundance of this domain has further enabled tasks such as virtual staining[91], in which models are trained to predict one type of image (e.g. a stained image) from another (e.g. a raw microscopy image). See Fig. 1b. Moving forward, AI algorithms that learn to perform diagnosis, prognosis, and theragnosis using digital pathology image archives and annotations readily available from electronic health records have the potential to transform the fields of pathology and oncology.

DERMATOLOGY
The key clinical tasks for DL in dermatology include lesion-specific differential diagnostics, finding concerning lesions amongst many benign lesions, and helping track lesion growth over time[92]. A series of works have demonstrated that CNNs can match the performance of board-certified dermatologists at classifying malignant skin lesions from benign ones[7,93,94]. These studies have sequentially tested increasing numbers of dermatologists (25–[7] 57–[93], 157–[94]), consistently demonstrating a sensitivity and specificity in classification that matches or even exceeds physician levels. These studies were largely restricted to the binary

classification task of discerning benign vs malignant cutaneous lesions, classifying either melanomas from nevi or carcinomas from seborrheic keratoses.

Recently, this line of work has expanded to encompass differential diagnostics across dozens of skin conditions[95], including non-neoplastic lesions such as rashes and genetic conditions, and incorporating non-visual metadata (e.g. patient demographics) as classifier inputs[96]. These works have been catalyzed by open-access image repositories and AI challenges that encourage teams to compete on predetermined benchmarks[97].

Incorporating these algorithms into clinical workflows would allow their utility to support other key tasks, including large-scale detection of malignancies on patients with many lesions, and tracking lesions across images in order to capture temporal features, such as growth and color changes. This area remains fairly unexplored, with initial works that jointly train CNNs to detect and track lesions[98].

OPHTHALMOLOGY

Ophthalmology, in recent years, has observed a significant uptick in AI efforts, with dozens of papers demonstrating clinical diagnostic and analytical capabilities that extend beyond current human capability[99–101]. The potential clinical impact is significant[102,103]—the portability of the machinery used to inspect the eye means that pop-up clinics and telemedicine could be used to distribute testing sites to underserved areas. The field depends largely on fundus imaging, and optical coherence tomography (OCT) to diagnose and manage patients.

CNNs can accurately diagnose a number of conditions. Diabetic retinopathy—a condition in which blood vessels in the eyes of diabetic patients "leak" and can lead to blindness—has been extensively studied. CNNs consistently demonstrate physician-level grading from fundus photographs[104–107], which has led to a recent US FDA-cleared system[108]. Similarly, they can diagnose or predict the progression of center-involved diabetic macular edema[109], age-related macular degeneration[107,110], glaucoma[107,111], manifest visual field loss[112], childhood blindness[113], and others.

The eyes contain a number of non-human-interpretable features, indicative of meaningful medical information, that CNNs can pick up on. Remarkably, it was shown that CNNs can classify a number of cardiovascular and diabetic risk factors from fundus photographs[114], including age, gender, smoking, hemoglobin-A1c, body-mass index, systolic blood pressure, and diastolic blood pressure. CNNs can also pick up signs of anemia[115] and chronic kidney disease[116] from fundus photographs. This presents an exciting opportunity for future AI studies predicting nonocular information from eye images. This could lead to a paradigm shift in care in which eye exams screen you for the presence of both ocular and nonocular disease—something currently limited for human physicians.

MEDICAL VIDEO

Surgical applications

The CV may provide significant utility in procedural fields such as surgery and endoscopy. Key clinical applications for deep learning include enhancing surgeon performance through real-time contextual awareness[117], skills assessments, and training. Early studies have begun pursuing these objectives, primarily in video-based robotic and laparoscopic surgery—a number of works propose methods for detecting surgical tools and actions[118–124]. Some studies analyze tool movement or other cues to assess surgeon skill[119,121,123,124], through established ratings such as the Global Operative Assessment of Laparoscopic Skills (GOALS) criteria for laparoscopic surgery[125]. Another line of work uses CV to recognize distinct phases of surgery during operations, towards developing
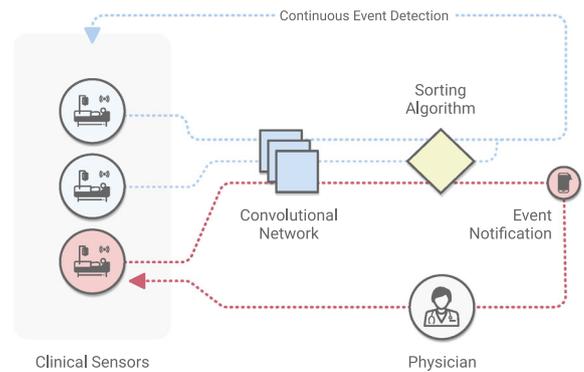

Fig. 3 Ambient intelligence. Computer vision coupled with sensors and video streams enables a number of safety applications in clinical and home settings, enabling healthcare providers to scale their ability to monitor patients. Primarily created using models for fine-grained activity recognition, applications may include patient monitoring in ICUs, proper hand hygiene and physical action protocols in hospitals and clinics, anomalous event detection, and others.

context-aware computer assistance systems[126,127]. CV is also starting to emerge in open surgery settings[128], of which there is a significant volume. The challenge here lies in the diversity of video capture viewpoints (e.g., head-mounted, side-view, and overhead cameras) and types of surgeries. For all types of surgical video, translating CV analysis to tools and applications that can improve patient outcomes is a natural next direction of research.

Human activity

CV can recognize human activity in physical spaces, such as hospitals and clinics, for a range of "ambient intelligence" applications. Ambient intelligence refers to a continuous, non-invasive awareness of activity in a physical space that can provide clinicians, nurses, and other healthcare workers with assistance such as patient monitoring, automated documentation, and monitoring for protocol compliance (Fig. 3). In hospitals, for example, early works have demonstrated CV-based ambient intelligence in intensive care units to monitor for safety-critical behaviors such as hand hygiene activity[32] and patient mobilization[8,129,130]. CV has also been developed for the emergency department, to transcribe procedures performed during the resuscitation of a patient[131], and for the operating room (OR), to recognize activities for workflow optimization[132]. At the hospital operations level, CV can be a scalable and detailed form of labor and resource measurement that improves resource allocation for optimal care[133].

Outside of hospitals, ambient intelligence can increase access to healthcare. For instance, it could enable at-risk seniors to live independently at home, by monitoring for safety and abnormalities in daily activities (e.g. detecting falls, which are particularly dangerous for the elderly[134,135]), assisted living, and physiological measurement. Similar work[136–138] has targeted broader categories of daily activity. Recognizing and computing long-term descriptive analytics of activities such as sleeping, walking, and sitting over time can detect clinically meaningful changes or anomalies[136]. To ensure patient privacy, researchers have developed CV algorithms that work with thermal video data[136]. Another application area of CV is assisted living or rehabilitation, such as continuous sign language recognition to assist people with communication difficulties[139], and monitoring of physiotherapy exercises for stroke rehabilitation[140]. CV also offers potential as a tool for remote physiological measurements. For instance, systems could use video[141] to analyze heart and breathing rates[141]. As telemedicine visits increase in frequency, CV could play a role in patient triaging,
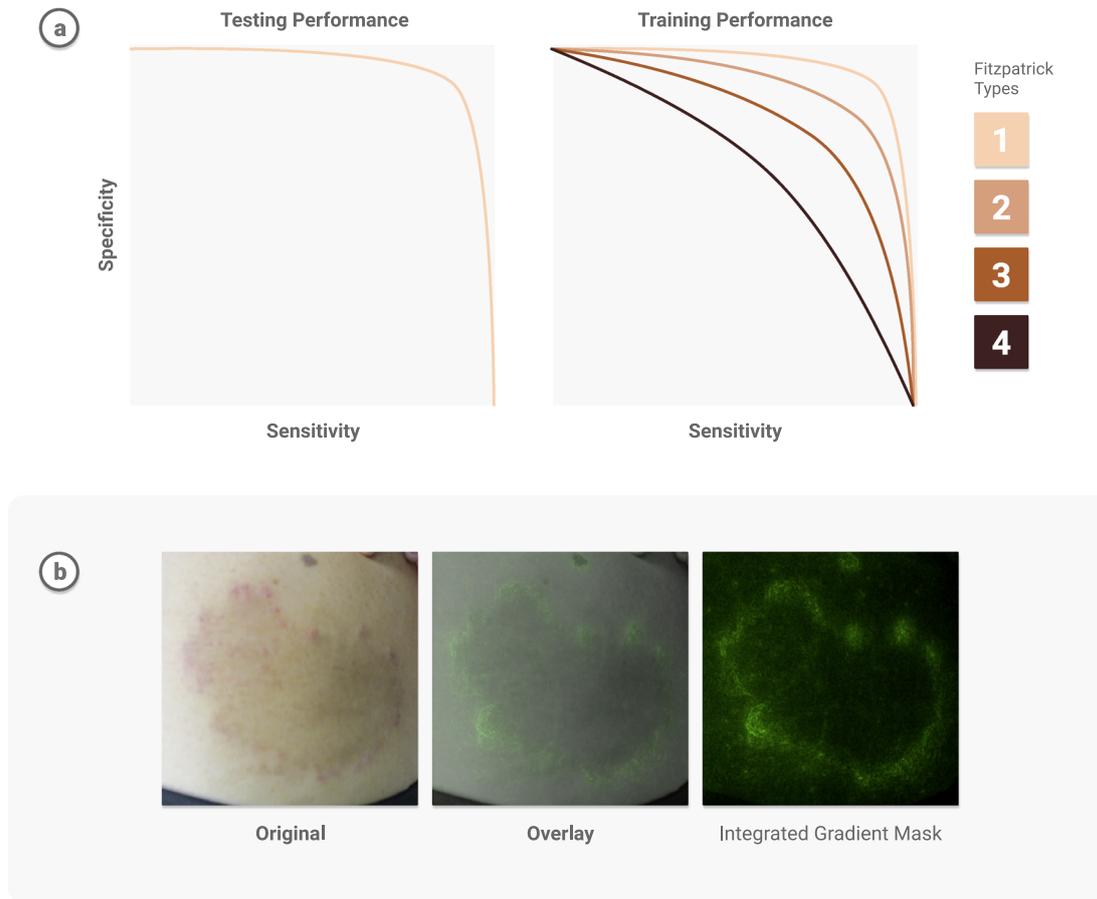
Fig. 4 Bias in deployment. a Example graphic of biased training data in dermatology. AIs trained primarily on lighter skin tones may not generalize as well when tested on darker skin[157]. Models require diverse training datasets for maximal generalizability (e.g.[95]). **b** Gradient Masks project the model's attention onto the original input image, allowing practitioners to visually confirm regions that most influence predictions. Panel was reproduced from ref. [95] with permission.

particularly in times of high demand such as the COVID-19 pandemic[142]. CV-based ambient intelligence technologies offer a wide range of opportunities for increased access to quality care.; However new ethical and legal questions will arise[143] in the design of these technologies.

CLINICAL DEPLOYMENT

As medical AI advances into the clinic[144], it will simultaneously have the power to do great good for society, and to potentially exacerbate long-standing inequalities and perpetuate errors in medicine. If done properly and ethically, medical AI can become a flywheel for more equitable care—the more it is used, the more data it acquires, the more accurate and general it becomes. The key is in understanding the data that the models are built on and the environment in which they are deployed. Here, we present four key considerations when applying ML technologies in healthcare: assessment of data, planning for model limitations, community participation, and trust building.

Data quality largely determines model quality; identifying inequities in the data and taking them into account will lead towards more equitable healthcare. Procuring the right datasets may depend on running human-in-the-loop programs or broad-reaching data collection techniques. There are a number of methods that aim to remove bias in data. Individual-level bias can be addressed via expert discussion[145] and labeling adjudication[146]. Population-level bias can be addressed via missing data supplements and distributional shifts. International multi-institutional

evaluation is a robust method to determine generalizability of models across diverse populations, medical equipment, resource settings, and practice patterns. In addition, using multi-task learning[147] to train models to perform a variety of tasks rather than one narrowly defined task, such as multi-cancer detection from histopathology images[148], makes them more generally useful and often more robust.

Transparent reporting can reveal potential weaknesses and help address model limitations. Guardrails to protect against possible worst-case scenarios—minority, dismissal, or automation bias—must be put in place. It is insufficient to report and be satisfied with strong performance measures on general datasets when delivering care for patients—there should be an understanding of the specific instances in which the model fails. One technique is to assess demographic performance in combination with saliency maps[149], to visualize what the model pays attention to, and check for potential biases. For instance, when using deep learning to develop a differential diagnosis for skin diseases[95], researchers examined the model performance based on Fitzpatrick skin types and other demographic information to determine patient types for which there were insufficient examples, and inform future data collection. Further, they used saliency masks to verify the model was informed by skin abnormalities and not skin type. See Fig. 4.

A known limitation of ML is its performance on out-of-distribution data–data samples that are unlike any seen during model training. Progress has been made on out-of-distribution detection[150] and developing confidence intervals to help detect anomalies. Additionally, methods are developing to understand
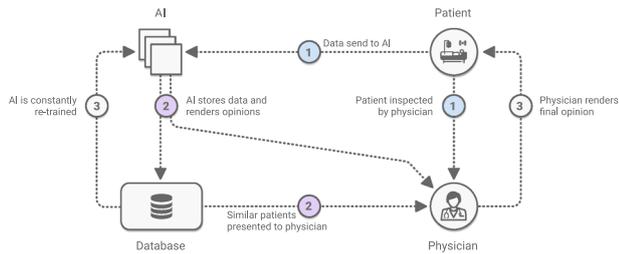
Fig. 5 Clinical Deployment. An example workflow showing the positive compounding effect of AI-enhanced workflows, and the resultant trust that can be built. AI predictions provide immediate value to physicians, and improve over time as bigger datasets are collected.

the uncertainty[151] around model outputs. This is especially critical when implementing patient-specific predictions that impact safety.

Community participation—from patients, physicians, computer scientists, and other relevant stakeholders—is paramount to successful deployment. This has helped identify structural drivers of racial bias in health diagnostics—particularly in discovering bias in datasets and identifying demographics for which models fail[152]. User-centered evaluations are a valuable tool in ensuring a system's usability and fit into the real world. What's the best way to present a model's output to facilitate clinical decision making? How should a mobile app system be deployed in resource-constrained environments, such as areas with intermittent connectivity? For example, when launching ML-powered diabetic retinopathy models in Thailand and India, researchers noticed that model performance was impacted by socioeconomic factors[38], and determined that where a model is most useful may *not* be where the model was generated. Ophthalmology models may need to be deployed in endocrinology care, as opposed to eye centers, due to access issues in the specific local environment. Another effective tool to build physician trust in AI results is side-by-side deployment of ML models with existing workflows (e.g manual grading[16]). See Fig. 5. Without question, AI models will require rigorous evaluation through clinical trials, to gauge safety and effectiveness. Excitingly, AI and CV can also help support clinical trials[153,154] through a number of applications—including patient selection, tumor tracking, adverse event detection, etc—creating an ecosystem in which AI can help design safe AI.

Trust for AI in healthcare is fundamental to its adoption[155] both by clinical teams and by patients. The foundation of clinical trust will come in large part from rigorous prospective trials that validate AI algorithms in real-world clinical environments. These environments incorporate human and social responses, which can be hard to predict and control, but for which AI technologies must account for. Whereas the randomness and human element of clinical environments are impossible to capture in retrospective studies, prospective trials that best reflect clinical practice will shift the conversation towards measurable benefits in real deployments. Here, AI interpretability will be paramount—predictive models will need the ability to describe *why* specific factors about the patient or environment lead them to their predictions.

In addition to clinical trust, patient trust—particularly around privacy concerns—must be earned. One significant area of need is next-generation regulations that account for advances in privacy-preserving techniques. ML typically does not require traditional identifiers to produce useful results, but there are meaningful signals in data that can be considered sensitive. To unlock insights from these sensitive data types, the evolution of privacy-preserving techniques must continue, and further advances need to be made in fields such as federated learning and federated analytics.

Each technological wave affords us a chance to reshape our future. In this case, artificial intelligence, deep learning, and computer vision represent an opportunity to make healthcare far more accessible, equitable, accurate, and inclusive than it has ever been.

DATA AVAILABILITY
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

REFERENCES
1. Szeliski, R. *Computer Vision: Algorithms and Applications* (Springer Science & Business Media, 2010).
2. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
3. Sanders, J. & Kandrot, E. CUDA by example: an introduction to general-purpose GPU programming. Addison-Wesley Professional; 2010 Jul 19. BibTeXEndNoteRefManRefWorks
4. Deng, J. et al. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009).
5. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29 (2019).
6. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
7. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
8. 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).
9. Russakovsky, O. et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
10. Krizhevsky, A., Sutskever, I. & Hinton, G. E. in *Advances in Neural Information Processing Systems* 25 (eds Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).
11. Sermanet, P. et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. Preprint at https://arxiv.org/abs/1312.6229 (2013).
12. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at https://arxiv.org/abs/1409.1556 (2014).
13. Szegedy, C. et al. Going deeper with convolutions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (CVPR) 1–9 (2015).
14. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
15. Gebru, T., Hoffman, J. & Fei-Fei, L. Fine-grained recognition in the wild: a multi-task domain adaptation approach. In *2017 IEEE International Conference on Computer Vision (ICCV)* 1358–1367 (IEEE, 2017).
16. Gulshan, V. et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in india. *JAMA Ophthalmol.* https://doi.org/10.1001/jamaophthalmol.2019.2004 (2014).
17. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* 234–241 (Springer, Cham, 2015).
18. Isensee, F. et al. nnU-Net: self-adapting framework for U-Net-based medical image segmentation. Preprint at https://arxiv.org/abs/1809.10486 (2018).
19. LeCun, Y. & Bengio, Y. in *The Handbook of Brain Theory and Neural Networks* 255–258 (MIT Press, 1998).
20. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. & Le, Q. V. AutoAugment: learning augmentation policies from data. Preprint at https://arxiv.org/abs/1805.09501 (2018).
21. Goodfellow, I. et al. Generative adversarial nets. In *Advances inneural information processing systems* 2672–2680 (2014).
22. Ørting, S. et al. A survey of Crowdsourcing in medical image analysis. Preprint at https://arxiv.org/abs/1902.09159 (2019).
23. Créquit, P., Mansouri, G., Benchoufi, M., Vivot, A. & Ravaud, P. Mapping of Crowdsourcing in health: systematic review. *J. Med. Internet Res.* **20**, e187 (2018).
24. Jing, L. & Tian, Y. in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE, 2020).

25. McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* 1273–1282 (PMLR, 2017).

26. Karpathy, A. & Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3128–3137 (IEEE, 2015).

27. Lv, D. et al. Research on the technology of LIDAR data processing. In *2017 First International Conference on Electronics Instrumentation Information Systems (EIIS)* 1–5 (IEEE, 2017).

28. Lillo, I., Niebles, J. C. & Soto, A. Sparse composition of body poses and atomic actions for human activity recognition in RGB-D videos. *Image Vis. Comput.* **59**, 63–75 (2017).

29. Haque, A. et al. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. In *Proceedings of the 2nd Machine Learning for Healthcare Conference*, **68**, 75–87 (PMLR, 2017).

30. Heilbron, F. C., Escorcia, V., Ghanem, B. & Niebles, J. C. ActivityNet: a large-scale video benchmark for human activity understanding. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 961–970 (IEEE, 2015).

31. Liu, Y. et al. Learning to describe scenes with programs. In *ICLR* (Open Access, 2019).

32. Singh, A. et al. Automatic detection of hand hygiene using computer vision-technology. *J. Am. Med. Inform. Assoc.* **27**, 1316–1320 (2020).

33. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).

34. Maron, O. & Lozano-Pérez, T. in *A Framework for Multiple-Instance Learning. in Advances in Neural Information Processing Systems* 10 (eds Jordan, M. I., Kearns, M. J. & Solla, S. A.) 570–576 (MIT Press, 1998).

35. Singh, S. P. et al. 3D Deep Learning On Medical Images: A Review. Sensors 20, https://doi.org/10.3390/s20185097 (2020).

36. Ouyang, D. et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).

37. Benjamens, S., Dhunnoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit. Med.* **3**, 118 (2020).

38. Beede, E. et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proc. 2020 CHI Conference on Human Factors in Computing Systems 1–12* (Association for Computing Machinery, 2020).

39. Viz.ai Granted Medicare New Technology Add-on Payment. PR Newswire https://www.prnewswire.com/news-releases/vizai-granted-medicare-new-technology-add-on-payment-301123603.html (2020).

40. Crowson, M. G. et al. A contemporary review of machine learning in otolaryngology-head and neck surgery. *Laryngoscope* **130**, 45–51 (2020).

41. Livingstone, D., Talai, A. S., Chau, J. & Forkert, N. D. Building an Otoscopic screening prototype tool using deep learning. *J. Otolaryngol. Head. Neck Surg.* **48**, 66 (2019).

42. Chen, P.-H. C. et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat. Med.* **25**, 1453–1457 (2019).

43. Gunčar, G. et al. An application of machine learning to haematological diagnosis. *Sci. Rep.* **8**, 411 (2018).

44. Alam, M. M. & Islam, M. T. Machine learning approach of automatic identification and counting of blood cells. *Health. Technol. Lett.* **6**, 103–108 (2019).

45. El Hajjar, A. & Rey, J.-F. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin. Med. J.* **133**, 326–334 (2020).

46. Horie, Y. et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest. Endosc.* **89**, 25–32 (2019).

47. Hirasawa, T. et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* **21**, 653–660 (2018).

48. Kubota, K., Kuroda, J., Yoshida, M., Ohta, K. & Kitajima, M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg. Endosc.* **26**, 1485–1489 (2012).

49. Itoh, T., Kawahira, H., Nakashima, H. & Yata, N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. *Endosc. Int Open* **6**, E139–E144 (2018).

50. He, J.-Y., Wu, X., Jiang, Y.-G., Peng, Q. & Jain, R. Hookworm detection in wireless capsule endoscopy images with deep learning. *IEEE Trans. Image Process.* **27**, 2379–2392 (2018).

51. Park, S.-M. et al. A mountable toilet system for personalized health monitoring via the analysis of excreta. *Nat. Biomed. Eng.* **4**, 624–635 (2020).

52. VerMilyea, M. et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum. Reprod.* **35**, 770–784 (2020).

53. Choy, G. et al. Current applications and future impact of machine learning in radiology. *Radiology* **288**, 318–328 (2018).

54. Saba, L. et al. The present and future of deep learning in radiology. *Eur. J. Radiol.* **114**, 14–24 (2019).

55. Mazurowski, M. A., Buda, M., Saha, A. & Bashir, M. R. Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. *J. Magn. Reson. Imaging* **49**, 939–954 (2019).

56. 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).

57. Irvin, J. et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proc. of the AAAI Conference on Artificial Intelligence* Vol. 33, 590–597 (2019).

58. Wang, X. et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervisedclassification and localization of common thorax diseases. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2097–2106 (2017).

59. Chilamkurthy, S. et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* **392**, 2388–2396 (2018).

60. Weston, A. D. et al. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* **290**, 669–679 (2019).

61. Ding, J., Li, A., Hu, Z. & Wang, L. in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017* 559–567 (Springer International Publishing, 2017).

62. Tan, L. K., Liew, Y. M., Lim, E. & McLaughlin, R. A. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Med. Image Anal.* **39**, 78–86 (2017).

63. Zhang, J. et al. Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection. Preprint at https://arxiv.org/abs/2003.12338 (2020).

64. Zhang, X., Feng, C., Wang, A., Yang, L. & Hao, Y. CT super-resolution using multiple dense residual block based GAN. *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, https://doi.org/10.1007/s11760-020-01790-5 (2020).

65. Papolos, A., Narula, J., Bavishi, C., Chaudhry, F. A. & Sengupta, P. P. U. S. Hospital use of echocardiography: insights from the nationwide inpatient sample. *J. Am. Coll. Cardiol.* **67**, 502–511 (2016).

66. HeartFlowNXT—HeartFlow Analysis of Coronary Blood Flow Using Coronary CT Angiography—Study Results—ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/results/NCT01757678.

67. Madani, A., Arnaout, R., Mofrad, M. & Arnaout, R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit. Med.* **1**, 6 (2018).

68. Zhang, J. et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation* **138**, 1623–1635 (2018).

69. Ghorbani, A. et al. Deep learning interpretation of echocardiograms. *NPJ Digit. Med.* **3**, 10 (2020).

70. Madani, A., Ong, J. R., Tibrewal, A. & Mofrad, M. R. K. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit. Med.* **1**, 59 (2018).

71. Perkins, C., Balma, D. & Garcia, R. Members of the Consensus Group & Susan G. Komen for the Cure. Why current breast pathology practices must be evaluated. A Susan G. Komen for the Cure white paper: June 2006. *Breast J.* **13**, 443–447 (2007).

72. Brimo, F., Schultz, L. & Epstein, J. I. The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. *J. Urol.* **184**, 126–130 (2010).

73. Elmore, J. G. et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* **313**, 1122–1132 (2015).

74. Evans, A. J. et al. US food and drug administration approval of whole slide imaging for primary diagnosis: a key milestone is reached and new questions are raised. *Arch. Pathol. Lab. Med.* **142**, 1383–1387 (2018).

75. Srinidhi, C. L., Ciga, O. & Martel, A. L. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis.* p. 101813 (2020).

76. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).

77. Cireşan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013* 411–418 (Springer Berlin Heidelberg, 2013).

78. Wang, H. et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *J. Med Imaging (Bellingham)* **1**, 034003 (2014).

79. Kashif, M. N., Ahmed Raza, S. E., Sirinukunwattana, K., Arif, M. & Rajpoot, N. Handcrafted features with convolutional neural networks for detection of tumor cells in histology images. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 1029–1032 (IEEE, 2016).

80. Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. Deep learning for identifying metastatic breast cancer. Preprint at https://arxiv.org/abs/1606.05718 (2016).

81. BenTaieb, A. & Hamarneh, G. in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* 460–468 (Springer International Publishing, 2016).

82. Chen, H. et al. DCAN: Deep contour-aware networks for object instance segmentation from histology images. *Med. Image Anal.* **36**, 135–146 (2017).

83. Xu, Y. et al. Gland instance segmentation using deep multichannel neural networks. *IEEE Trans. Biomed. Eng.* **64**, 2901–2912 (2017).

84. Litjens, G. et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **6**, 26286 (2016).

85. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).

86. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).

87. Mobadersany, P. et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl Acad. Sci. U. S. A.* **115**, E2970–E2979 (2018).

88. Courtiol, P. et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* **25**, 1519–1525 (2019).

89. Rawat, R. R. et al. Deep learned tissue 'fingerprints' classify breast cancers by ER/PR/Her2 status from H&E images. *Sci. Rep.* **10**, 7275 (2020).

90. Dietterich, T. G., Lathrop, R. H. & Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* **89**, 31–71 (1997).

91. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803.e19 (2018).

92. Esteva, A. & Topol, E. Can skin cancer diagnosis be transformed by AI? *Lancet* **394**, 1795 (2019).

93. Haenssle, H. A. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **29**, 1836–1842 (2018).

94. Brinker, T. J. et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur. J. Cancer* **113**, 47–54 (2019).

95. Liu, Y. et al. A deep learning system for differential diagnosis of skin diseases. *Nat. Med.* **26**, 900–908 (2020).

96. Yap, J., Yolland, W. & Tschandl, P. Multimodal skin lesion classification using deep learning. *Exp. Dermatol.* **27**, 1261–1267 (2018).

97. Marchetti, M. A. et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J. Am. Acad. Dermatol.* **78**, 270–277 (2018).

98. Li, Y. et al. Skin cancer detection and tracking using data synthesis and deep learning. Preprint at https://arxiv.org/abs/1612.01074 (2016).

99. Ting, D. S. W. et al. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **103**, 167–175 (2019).

100. Keane, P. A. & Topol, E. J. With an eye to AI and autonomous diagnosis. *NPJ Digit. Med.* **1**, 40 (2018).

101. Keane, P. & Topol, E. Reinventing the eye exam. *Lancet* **394**, 2141 (2019).

102. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).

103. Kern, C. et al. Implementation of a cloud-based referral platform in ophthalmology: making telemedicine services a reality in eye care. *Br. J. Ophthalmol.* **104**, 312–317 (2020).

104. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).

105. Raumviboonsuk, P. et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit Med.* **2**, 25 (2019).

106. Abràmoff, M. D. et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest. Ophthalmol. Vis. Sci.* **57**, 5200–5206 (2016).

107. Ting, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* **318**, 2211–2223 (2017).

108. Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N. & Folk, J. C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit. Med.* **1**, 39 (2018).

109. Varadarajan, A. V. et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nat. Commun.* **11**, 130 (2020).

110. Yim, J. et al. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat. Med.* **26**, 892–899 (2020).

111. Li, Z. et al. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* **125**, 1199–1206 (2018).

112. Yousefi, S. et al. Detection of longitudinal visual field progression in glaucoma using machine learning. *Am. J. Ophthalmol.* **193**, 71–79 (2018).

113. Brown, J. M. et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol.* **136**, 803–810 (2018).

114. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).

115. Mitani, A. et al. Detection of anaemia from retinal fundus images via deep learning. *Nat. Biomed. Eng.* **4**, 18–27 (2020).

116. Sabanayagam, C. et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digital Health* **2**, e295–e302 (2020).

117. Maier-Hein, L. et al. Surgical data science for next-generation interventions. *Nat. Biomed. Eng.* **1**, 691–696 (2017).

118. García-Peraza-Herrera, L. C. et al. ToolNet: Holistically-nested real-time segmentation of robotic surgical tools. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 5717–5722 (IEEE, 2017).

119. Zia, A., Sharma, Y., Bettadapura, V., Sarin, E. L. & Essa, I. Video and accelerometer-based motion analysis for automated surgical skills assessment. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 443–455 (2018).

120. Sarikaya, D., Corso, J. J. & Guru, K. A. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Trans. Med. Imaging* **36**, 1542–1549 (2017).

121. Jin, A. et al. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* 691–699 (IEEE, 2018).

122. Twinanda, A. P. et al. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans. Med. Imaging* **36**, 86–97 (2017).

123. Lin, H. C., Shafran, I., Yuh, D. & Hager, G. D. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Comput. Aided Surg.* **11**, 220–230 (2006).

124. 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).

125. Vassiliou, M. C. et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am. J. Surg.* **190**, 107–113 (2005).

126. Jin, Y. et al. SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. *IEEE Trans. Med. Imaging* **37**, 1114–1126 (2018).

127. Padoy, N. et al. Statistical modeling and recognition of surgical workflow. *Med. Image Anal.* **16**, 632–641 (2012).

128. Azari, D. P. et al. Modeling surgical technical skill using expert assessment for automated computer rating. *Ann. Surg.* **269**, 574–581 (2019).

129. Ma, A. J. et al. Measuring patient mobility in the ICU using a novel noninvasive sensor. *Crit. Care Med.* **45**, 630–636 (2017).

130. Davoudi, A. et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci. Rep.* **9**, 8020 (2019).

131. Chakraborty, I., Elgammal, A. & Burd, R. S. Video based activity recognition in trauma resuscitation. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* 1–8 (IEEE, 2013).

132. Twinanda, A. P., Alkan, E. O., Gangi, A., de Mathelin, M. & Padoy, N. Data-driven spatio-temporal RGBD feature encoding for action recognition in operating rooms. *Int. J. Comput. Assist. Radiol. Surg.* **10**, 737–747 (2015).

133. Kaplan, R. S. & Porter, M. E. How to solve the cost crisis in health care. *Harv. Bus. Rev.* **89**, 46–52 (2011). 54, 56–61 passim.

134. Wang, S., Chen, L., Zhou, Z., Sun, X. & Dong, J. Human fall detection in surveillance video based on PCANet. *Multimed. Tools Appl.* **75**, 11603–11613 (2016).

135. Núñez-Marcos, A., Azkune, G. & Arganda-Carreras, I. Vision-Based Fall Detection with Convolutional Neural Networks. In *Proc. International Wireless Communications and Mobile Computing Conference 2017* (ACM, 2017).

136. Luo, Z. et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. In *Machine Learning for Healthcare (MLHC)* 2 (JMLR, 2018).

137. Zhang, C. & Tian, Y. RGB-D camera-based daily living activity recognition. *J. Comput. Vis. image Process.* **2**, 12 (2012).

138. Pirsiavash, H. & Ramanan, D. Detecting activities of daily living in first-person camera views. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* 2847–2854 (IEEE, 2012).

139. Kishore, P. V. V., Prasad, M. V. D., Kumar, D. A. & Sastry, A. S. C. S. Optical flow hand tracking and active contour hand shape features for continuous sign

language recognition with artificial neural networks. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)* 346–351 (IEEE, 2016).

140. Webster, D. & Celik, O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *J. Neuroeng. Rehabil.* **11**, 108 (2014).

141. Chen, W. & McDuff, D. Deepphys: video-based physiological measurement using convolutional attention networks. In *Proc. European Conference on Computer Vision (ECCV)* 349–365 (Springer Science+Business Media, 2018).

142. Moazzami, B., Razavi-Khorasani, N., Dooghaie Moghadam, A., Farokhi, E. & Rezaei, N. COVID-19 and telemedicine: Immediate action required for maintaining healthcare providers well-being. *J. Clin. Virol.* **126**, 104345 (2020).

143. Gerke, S., Yeung, S. & Cohen, I. G. Ethical and legal aspects of ambient intelligence in hospitals. *JAMA* https://doi.org/10.1001/jama.2019.21699 (2020).

144. Young, A. T., Xiong, M., Pfau, J., Keiser, M. J. & Wei, M. L. Artificial intelligence in dermatology: a primer. *J. Invest. Dermatol.* **140**, 1504–1512 (2020).

145. Schaekermann, M., Cai, C. J., Huang, A. E. & Sayres, R. Expert discussions improve comprehension of difficult cases in medical image assessment. In *Proc. 2020 CHI Conference on Human Factors in Computing Systems* 1–13 (Association for Computing Machinery, 2020).

146. Schaekermann, M. et al. Remote tool-based adjudication for grading diabetic retinopathy. *Transl. Vis. Sci. Technol.* **8**, 40 (2019).

147. Caruana, R. Multitask learning. *Mach. Learn.* **28**, 41–75 (1997).

148. Wulczyn, E. et al. Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLoS ONE* **15**, e0233678 (2020).

149. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep inside convolutional networks: visualising image classification models and saliency maps. Preprint at https://arxiv.org/abs/1312.6034 (2013).

150. Ren, J. et al. in *Advances in Neural Information Processing Systems* 32 (eds Wallach, H. et al.) 14707–14718 (Curran Associates, Inc., 2019).

151. Dusenberry, M. W. et al. Analyzing the role of model uncertainty for electronic health records. In *Proc. ACM Conference on Health, Inference, and Learning* 204–213 (Association for Computing Machinery, 2020).

152. 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).

153. Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ* **370**, m3164 (2020).

154. Rivera, S. C. et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ* **370**, m3210 (2020).

155. Asan, O., Bayrak, A. E. & Choudhury, A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J. Med. Internet Res.* **22**, e15154 (2020).

156. McKinney, S. M. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).

157. Kamulegeya, L. H. et al. Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning. https://doi.org/10.1101/826057 (2019).

AUTHOR CONTRIBUTIONS
A.E. organized the authors, synthesized the writing, and led the abstract, introduction, computer vision, dermatology, and ophthalmology sections. S.Y. led the medical video section. K.C. led the clinical deployment section. N.N. contributed the pathology section, Ali Madani contributed the cardiology section, Ali Mottaghi contributed to the sections within the medical video, and E.T. and J.D. contributed to the clinical deployment section. Y.L. significantly contributed to the figures, and writing style. All authors contributed to the overall writing and storyline. E.T., J.D., and R.S. oversaw and advised the work.

COMPETING INTERESTS
A.E., N.N., Ali Madani, and R.S. are or were employees of Salesforce.com and own Salesforce stock. K.C., Y.L., and J.D. are employees of Google, L.L.C. and own Alphabet stock. S.Y., Ali Mottaghi and E.T. have no competing interests to declare.

ADDITIONAL INFORMATION
Correspondence and requests for materials should be addressed to A.E.

Reprints and permission information is available at http://www.nature.com/reprints

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