## Digital Oncology 2

# Big data and machine learning algorithms for health-care delivery

*Kee Yuan Ngiam, Ing Wei Khor*

**Analysis of big data by machine learning offers considerable advantages for assimilation and evaluation of large amounts of complex health-care data. However, to effectively use machine learning tools in health care, several limitations must be addressed and key issues considered, such as its clinical implementation and ethics in health-care delivery. Advantages of machine learning include flexibility and scalability compared with traditional biostatistical methods, which makes it deployable for many tasks, such as risk stratification, diagnosis and classification, and survival predictions. Another advantage of machine learning algorithms is the ability to analyse diverse data types (eg, demographic data, laboratory findings, imaging data, and doctors' free-text notes) and incorporate them into predictions for disease risk, diagnosis, prognosis, and appropriate treatments. Despite these advantages, the application of machine learning in health-care delivery also presents unique challenges that require data pre-processing, model training, and refinement of the system with respect to the actual clinical problem. Also crucial are ethical considerations, which include medico-legal implications, doctors' understanding of machine learning tools, and data privacy and security. In this Review, we discuss some of the benefits and challenges of big data and machine learning in health care.**

*Lancet Oncol* 2019; 20: e262–73

This is the second in a **Series** of two papers about Digital Oncology

**Department of Surgery** (K Y Ngiam MBBS) **and Department of Medicine** (I W Khor PhD), **Yong Loo Lin School of Medicine, National University of Singapore, Singapore; Division of General Surgery (Thyroid and Endocrine Surgery), University Surgical Cluster, National University Hospital, Singapore** (K Y Ngiam)**; and National University Health System Corporate Office, Singapore** (K Y Ngiam)

Correspondence to: Dr Kee Yuan Ngiam, National University Health System Corporate Office, Singapore 119228 **kee_yuan_ngiam@nuhs.edu.sg**

## Introduction

Machine learning is a type of artificial intelligence (AI) that encompasses algorithmic methods that enable machines to solve problems without specific computer programing.[1] The term AI is used rather loosely in the literature and popular media to describe a broad range of promising applications, such as self-driving vehicles, digital personal assistants, and personalisation of consumer products. In this Review, we will refer to machine learning as the specific class of tools used for processing data and how these apply to a health-care context. Although much excitement surrounds the use of AI in health care and other fields, the promise of self-learning, continuously advancing machine learning algorithms needs to be tempered against the challenges of implementing such tools in routine clinical practice. To frame these challenges correctly, defining the intent and scope of deploying such tools is an important step before implementation. This Review will cover the key issues and limitations surrounding the application of machine learning in health-care delivery.

Broadly, there are three points that should be considered: a technical appreciation of machine learning tools (including an understanding of their limitations and how to interpret their findings) an ethical, regulatory, and legal framework for the safe use of these tools in clinical practice, and a governance framework and platform to enable equitable use of data.

With regard to the technical considerations, machine learning is essentially an interplay between large datasets with a specific class of machine learning methods called deep neural networks or deep learning. Popularised in 2012, these neural networks can be trained to be highly accurate in finding complex patterns within big data.[2] These deep learning networks can also be retrained with population-specific datasets, and be used for multiple health-care applications.[2–7] Health-care data are notorious for being voluminous, messy, and complex. For data to be useful, they must be appropriately mapped and pre-processed before they can be used to train machine learning methods. This step is fundamental in machine learning model building, because the accuracy of the model is highly dependent on the reliability of the data in terms of its reflection of clinical reality. For example, if a wrong medication is inadvertently included as part of the dataset used to train a machine learning model to suggest treatments for a particular condition, the model might erroneously suggest this medication for the condition, leading to disastrous consequences.

Just like any tool, a machine learning algorithm must be appropriately engineered to be truly effective. In a health-care context, the clinical problem must be both the driver and the reference point for machine learning applications. With the clinical problem as the focal point, the ability of machine learning to assimilate and analyse large and diverse datasets comprising different types of clinical data makes it an invaluable aid to clinicians in making decisions for the care of their patients. Using this tool, clinicians can take into consideration more pieces of evidence than they could otherwise process and remember of their own accord.[5,6,8]

Another important consideration is the ethics surrounding the use of machine learning in health care. Clear guidelines that have been developed with clinicians have lagged behind the advances in machine learning. However, guidelines have emerged, such as Singapore's Model Artificial Intelligence Governance Framework,[9] which guides private sector organisations on how to use

AI ethically. In agreement with many observers, we believe that the most ethically feasible scenario involves the use of AI to augment the capability of human doctors, rather than replace them.[10,11] Machines cannot reproduce the emotional virtues that human doctors are capable of, such as empathy, compassion, and care.[12] AI can take over tasks that are more routine or standardised, to which they are very well suited for, freeing up time so that doctors can spend more time on tasks that need human judgment, intuition, or empathy.[5,13] This scenario allows for a clearer assignment of medical liability (to the doctor) and circumvents issues that could arise when the machine produces different results or performs better than the doctor or when the doctor does not consider the machine's predictions in clinical decision making. Other issues surrounding data privacy, security, and control require the implementation of strong data anonymisation and security measures to prevent data breaches.

For a machine learning tool to be effective in delivery of health care, it should ideally be integrated into a single platform. This arrangement affords central control over functions such as data pre-processing, data governance, regulatory requirements, and operational interaction with existing electronic health record systems. The juxtaposition of these functions is essential for operational deployment of machine learning tools in clinical practice. Several companies have attempted to develop so-called AI platforms to house suites of modular machine learning tools for development in a clinical environment. Some institutions have proprietary platforms that are integrated with electronic health records and incorporate unique data governance structures that take advantage of cloud-based machine learning platforms. These platforms are described in the Examples of machine learning platforms section.

## Big data and the promise of population health advancements

The advent of big data and the widespread use of electronic health records for patients have enabled us to pursue solutions to population health issues previously thought impossible. Instead of extrapolating from the data obtained from a small number of samples to make inferences about a population, we can now use clinical data at the population level to provide a real-world picture. Analysing the actual data from large numbers of people is a fundamental change from classical biostatistics, which focuses on reducing the effects of all kinds of bias due to study design. Although randomised controlled trials remain the gold standard for establishing the effectiveness of a particular drug, observing the effectiveness of the drug at the population level, which includes real-world factors such as drug compliance, provides a better model of the true effectiveness of the drug.[14] With the availability of population-level electronic records spanning several decades, study of the longitudinal effects of treatments at the population level is possible, which is especially important for the evaluation of treatment consequences that develop or evolve over time. This kind of big data, collected over long periods of time, can be used to build AI models that provide predictions for future events on the basis of the statistical weight of historical correlations.

Another powerful dimension to the pooling of electronic data is the linkage of phenotypic data to research data. Increasingly, patient registries and databases are being linked to genotypic and other research data in one location and made accessible to clinicians from different specialties. However, this integration of datasets is not a trivial issue, because it exposes issues regarding data standards, governance, anonymisation, equitable data sharing, and compliance with data protection laws. If implemented well, the linkage of previously siloed datasets exposes new possibilities for discovering genetic, biological, and clinical associations that might explain disease pathogenesis and progression. It also enables the evaluation of the effects of patient and disease characteristics on outcomes and health-care usage.[6] Some of these applications are now possible because of unique machine learning features, such as the flexibility to combine different neural network algorithms, resulting in complex and powerful deep learning models. This deep learning infrastructure also confers flexibility and scalability on models, enabling the same set of algorithms to be iteratively trained to address multiple clinical problems found in the depth and vastness of the underlying datasets.[7]

These scalable tools could more accurately stratify patients into subgroups based on their predicted risk of disease. One application of such predictive analytics is the identification of a subgroup of patients with a higher risk of hospital admission, who are thus likely to account for the majority of health-care costs, and implementation of timely pre-emptive interventions based on such predictions.[3] If successfully validated and implemented, these risk stratification tools might dramatically reduce the cost and morbidity associated with readmissions that could have been averted.

AI platforms also present a substantial opportunity to improve patient-facing services through pooling of data sources and centralisation of services. An example would be telehealth, where multisource information from telephone, text, and video consults might be aggregated into patient records, which can then be analysed by machine learning tools. Much of the excitement surrounding telehealth is focused on providing convenience and accessibility to patients through modern communications technology. However, two studies[15,16] have shown that such practices do not always demonstrate cost-effectiveness or improvements in health-related (eg, morbidity, mortality, quality of life, and patient satisfaction), process-related (eg, quality of care, professional practice, adherence to recommended practice, and professional satisfaction), or resource-use outcomes that were promised, largely because of the dependence on human operators for tasks that could have otherwise been done by machines, such as scheduling appointments and filling prescriptions. With AI chatbots using natural language processing, it is possible to do these tasks with an intuitive machine that can not only address standard telephone call requests, but also complete advanced tasks, such as triaging, preventive screening, and even image diagnosis.[17] These tools promise to transform telehealth into an effective means of scaling health-care provision at the community level. However, few chatbots have been evaluated for health-care applications in randomised controlled trials.[18,19]

An important caveat is that these promises of population-level health advancement are contingent on the integrity of the data. A major disadvantage of retrospective clinical data is their high variability due to the absence of standardised methods of data collection. This variability results in the noisiness and so-called missingness of follow-up data. Fortunately, robust methods exist[20] to adjust for the bias and irregularity in sufficiently large datasets, which render the data still usable for AI model building.

## Data requirements and approaches

### Deep learning and big data

Deep learning is uniquely suited to handling big data because of its capacity to assimilate big datasets and make sense of the complex relationships between variables in a flexible, trainable manner.[7] In deep neural networks, multiple sequential layers of intermediate variables connect input features and outputs, so that the outputs of one layer serve as inputs of the next layer.[2,7] This structure facilitates the analysis of high-dimensional data, defined as data involving more than 200 variables. In contrast, traditional statistical methods, such as linear regression, involve only one input-output layer and can accommodate relatively small amounts of variation.

A specific type of computing hardware called graphics processing units (GPUs) are optimised to process these multilayer neural networks. With the widespread availability of large amounts of RAM and computers equipped with GPUs since 2012, deep neural networks can now be deployed in complex predictive models for various clinical purposes, including readmission prediction, imaging analysis, and drug discovery.[3,21–23]

Despite its advantages, deep learning is not a panacea to all modelling tasks. Numerous machine learning methods, such as support vector machines, decision trees, and Bayesian networks, might adequately address many health-care tasks. With its capacity for backpropagation and the ability to stack multiple layers, deep learning is distinct from other machine learning methods and is suited to handling data with many variables. For example. as many as 69 000 ICD-10-CM diagnostic variables and more than 24 000 medication types exist.

### Data pre-processing

For big data to be used in solving clinical problems, the clinical data must first be carefully labelled and curated. Medical data are diverse and can take the form of doctor's notes (in long-form text), clinical laboratory reports, clinical images, and information from medical devices. The labels used for these data must accurately reflect clinical reality because they will be used to train the machine learning system. Any inaccuracy in labelling will severely limit the accuracy attainable by the machine learning algorithm, regardless of the effort invested in improving the algorithm. Thus, to discern data accurately, data scientists and clinicians typically invest substantial amounts of time and effort in ensuring the reliability of their data before they embark on model building.[24]

Data labelling can be a challenging process for several reasons. The expansion of diagnostic terminology from one version of diagnostic criteria to the next (eg, the ninth and tenth editions of the International Classification of Diseases) creates ambiguity in diagnostic label transitions, especially in longitudinal datasets. The long-form text that doctors use to describe a specific condition and its symptoms can be as varied as the doctors themselves. In this scenario, the data need to be meticulously relabelled by considering the context of the clinical history. For example, the term myocardial infarction has different clinical implications when it occurs in the section of the electronic health record describing the patient's personal medical history compared with when it occurs in the family history section of the record.[25]

The next part of training a machine learning system for clinical applications centres on data curation, a process of reclassifying data into clinically or logically relevant subgroups that might improve the predictive accuracy of the machine learning tool for the intended clinical problem. This step requires substantial clinical understanding of the data, the nature of the problem, and the performance limitations of the machine learning methods used. For example, doctors might need to manually reclassify the systems-based Diagnosis Related Groups codes into concept-level severity classes that more appropriately express the effect of diagnostic severity on clinical outcomes. To facilitate logical reasoning in machine learning, the format and semantics of data should also be consistent across clinical systems.[25] In this aspect, ontologies, or sets of terminologies that show the relationships between concepts, can be very helpful. Ontologies, such as the Unified Medical Language System, provide standardised names, synonyms, and cross-references across more than 100 different clinical terminologies and coding systems.[26] Ontologies are also available for different types of clinical data, including diagnoses, procedures, medications, and laboratory findings.

Although neural networks do not require ideal datasets, they do require data points to be at regular intervals and for the data to be complete (ie, no missing data points).

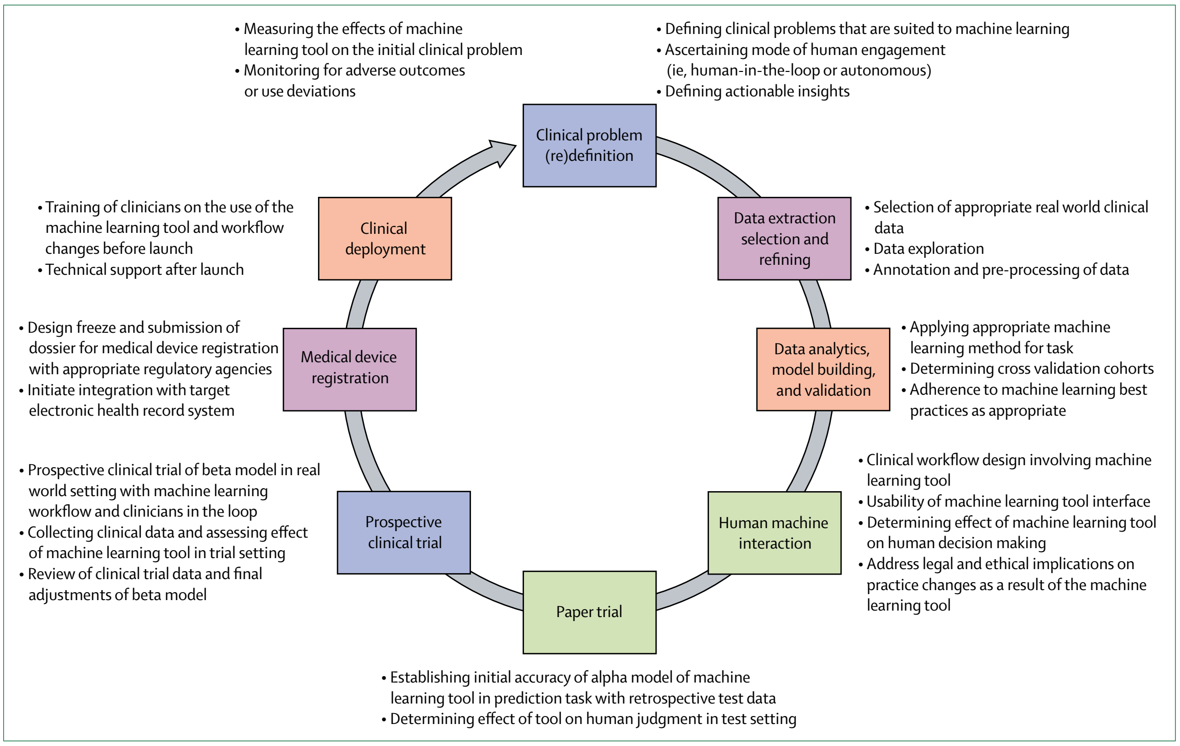


***Figure 1:*** **Training, clinical trial evaluation, and clinical implementation of machine learning algorithms for health-care applications**

Data that do not meet these requirements must be standardised during the data pre-processing phase. For example, if data points occur irregularly (as with much of health-care data), they need to be adjusted to fit the predetermined time points—eg, by systematically imputing data at a number of days before or after the nearest predetermined time point to that time point. If data points are missing, they need to be statistically imputed if certain criteria are met. Alternatively, the variable can be dropped if data points are insufficient or the data are not variable, because it might not affect the result. Furthermore, structured data points with reference ranges (eg, laboratory results) must be normalised according to the institution-specific reference ranges for the neural network model to understand the significance of the results.

## Neural networks

After appropriate pre-processing of the data, the next step is to select the appropriate neural network for the intended clinical task. Numerous variants of neural networks exist, such as convolutional neural networks and recurrent neural networks.

A convolutional neural network typically involves several steps, starting with the application of a convolution filter to pre-processed data.[27] The filter enables the network to identify a particular shape in an image or word in a text document by scanning the input data and calculating a value that represents a summary of the features in the input.[7,23] The results from this step are then processed to introduce non-linearity to the network, because most real-world data are non-linear.[28] Further processing is then done to ensure that the probabilities of all predicted outputs (labels) add up to one. The aforementioned process produces only one prediction (the most probable) for a given set of inputs. Another important feature of neural networks is backpropagation, which involves training the model from a known end-result (output to input) through small, iterative adjustments to ensure that the labels or predictions produced more accurately describe the end-result. A potential pitfall of backpropagation is that the model becomes too specific for a particular clinical scenario (overfitting), in which case noise can be introduced into the datasets to teach the model more generic features.[29]

A recurrent neural network has the features of a convolutional neural network, but also includes the memory of previous inputs. The recurrent neural network considers both the current input and previous inputs when making a decision.

Convolutional neural networks have been shown to do well in the area of image recognition and recurrent neural networks in natural language processing.[7,30] However, convolutional neural networks and recurrent neural networks are flexible and can be adapted to work with electronic health record data either singly or in combination in areas such as speech recognition. A detailed examination of the iterative process of experimenting with various model architectures is beyond the remit of this Review. However, to effectively apply neural networks to clinical problems, a machine learning framework and deep learning best practices must be in place to ensure that the resultant deep neural network architecture is suitable for big data processing and generalisable to multiple tasks.[31] For example, data scientists should use a systematic and empirical approach to engineer neural network architectures that are commensurate to the task, use an appropriate number of neural network layers, and adjust the weights (hyperparameter tuning) to maximise the chances of improving model outcomes.

## Training machine learning for clinical applications

Training machine learning tools for clinical application is vastly different from training research machine learning tools. Most clinical machine learning tools are based on supervised learning methods, in which data are classified into predetermined categories. The bar for accuracy and clinical efficacy of clinical machine learning tools approaches that of regulated medical devices. The US Food & Drug Administration (FDA) issued guidance on software as a medical device in 2017, explaining risk stratification and the analytical and clinical validation required of AI tools in health care.[32,33]

Unlike medical devices, a unique feature of AI tools is their ability to be continuously improved with new data. This process is called incremental learning, in which outcomes data from a trained AI system are incorporated into a closed data feedback loop and used to refine the predictive accuracies of the system through iterative retraining of the model.[34,35] This feature distinguishes trainable neural networks from immutable scoring systems or standardised software. Figure 1 shows the training process, clinical trial evaluation, and its implementation in clinical settings for machine learning algorithms in health-care delivery.

An example of a machine learning tool that has been finely tuned to the actual clinical problem is IDx-DR, the first FDA-approved machine learning tool to provide unaided screening decisions (table 1).[36,51] The software program uses a machine learning algorithm to analyse the retinal images from a patient, producing one of two possible screening results: positive (more than mild diabetic retinopathy) and negative (mild diabetic retinopathy or lower). This diabetic retinopathy diagnostic tool was the first to be tested in the clinic in a prospective trial, exhibiting high sensitivity (87%) and specificity (91%) for diabetic retinopathy when used in primary care clinics, which might not routinely perform eye care. Although the IDx-DR platform does not require a clinician to confirm the screening results, its purpose is still to provide a recommendation to the clinician, who ultimately makes the treatment decision.

Several reviews[11,37,40] have discussed IDx-DR and other machine learning tools that are at different stages of clinical development (table 1; table 2). Besides binary platforms, such as IDx-DR, some machine learning systems provide multilabel read-outs. For example, a deep neural network analysing ultrasound images was able to grade prostate cancer tumours into three categories (aggressive, less clinically significant, and benign) at a sensitivity and specificity greater than 70%.[47] Another deep neural network analysed tumour histology images and genomic markers to produce more accurate cancer survival predictions compared with conventional histology and genomic marker tests.[48] The algorithm considered intratumoural heterogeneity during model training by randomly sampling fields within regions of interest in histological images.

However, as noted previously, machine learning has limitations that can lead to inaccurate predictions in some clinical scenarios. A case in point is the problematic application of the IBM (Armonk, NY) supercomputer Watson in oncology (table 1). According to a news report based on viewing internal IBM documents, the Watson for Oncology application has produced many “unsafe and incorrect” treatment recommendations that contradicted clinical practice guidelines.[49] The main issues with this platform appear to be that it was trained on data from a relatively small number of patients with cancer, rather than pooled patient datasets with appropriate controls, and that the recommendations were based on a few doctors’ advice instead of established treatment guidelines. This highlights the importance of curating real-world clinical data when it is used to train machine learning algorithms.

## Human–machine interactions

Another important aspect of the development of machine learning tools is their effect on human decision making. The key premise of developing machine learning technologies is the augmentation of human performance in health-care workflows that are repetitive, mundane, or simply impossible for humans to accurately estimate (eg, mortality risk). Many such tools can only function as clinical decision support systems, at least in part because of the substantial ethical and medico-legal considerations surrounding the use of these tools autonomously.

Because even the best models can be misinterpreted by doctors, it is imperative that machine learning models be tested to determine the effect of human–machine interactions in real-life clinical situations, not ideal textbook scenarios. For example, models can be trained using data produced by clinical activities such as diagnosis, treatment selection, and treatment monitoring.[52] Similar to the training of junior doctors, a clinical machine learning tool is best trained by incorporating real-world medical data into the disease model, then tuned by medical experts to improve its accuracy in predicting real cases.[6] Unlike junior doctors, a machine learning system can incorporate vast amounts of historical clinical data, as well as the latest peer-reviewed medical guidelines into its model.

| | Type of data used or clinical application | Comments | External validation | Validated with humans | Validated against humans | US Food & Drug Administration approval |
|---|---|---|---|---|---|---|
| **Eye diseases** | | | | | | |
| Deep neural network for diagnosing diabetic retinopathy (IDx-DR, IDx Technologies, Coralville, IA)[36] | Imaging data (retinal images) | Provides a binary read-out (ie, yes or no) for more-than-mild diabetic retinopathy. Although it does not require clinician confirmation of results, it is designed to be used together with clinicians | Yes | Yes | Yes | Yes |
| Random forest algorithm for predicting myopia[37] | Electronic health record data | Algorithm predicted development of adult myopia in school children up to 8 years before onset, at an accuracy of 85% to 99% (area under the curve). A drawback is that the refraction measurements were taken by several different optometrists | Yes | Yes | No | No |
| **Cardiac abnormalities** | | | | | | |
| Cloud-based deep neural network algorithm (Cardio DL, Arterys, San Francisco, CA) for diagnosing cardiac anomalies[38] | Imaging data (MRI of heart ventricles) | First cloud-based artificial intelligence tool to be approved by the US Food & Drug Administration. Shown to assess heart ventricular function and blood flow as well as radiologists. Uses a system that removes patient identifiers from images at point of data collection (usually hospitals), so only de-identified data is stored. Lack of peer-reviewed studies makes it difficult to fully evaluate the algorithm | Yes | Yes | Yes | Yes |
| **Fractures** | | | | | | |
| OsteoDetect (Imagen, Cambridge, UK) for detecting distal radius fractures in the wrist[39] | Imaging data (X-ray) | Clinicians using the algorithm showed better sensitivity, specificity, and positive and negative predictive values for wrist fracture diagnosis compared with clinicians without the algorithm. A limitation is that the supporting studies were both retrospective | Yes | Yes | Yes | Yes |
| **Pulmonary diseases** | | | | | | |
| Deep neural network algorithm (CheXNeXt) for detecting 14 diseases, including pneumonia, pulmonary masses, and pleural effusion[40] | Imaging data (chest radiographs) | Validated using subsets of the ChestX-ray8 dataset from the National Institutes of Health, which contains over 100 000 chest radiographs. When retrospectively compared with 9 radiologists' performance, algorithm performed as well for 10 diseases, better on 1, and worse on 3, in a much shorter time (1·5 min *vs* 240 min for pathologists). However, the study was retrospective, did not replicate a real-world clinical setting, and involved only one site | No | No | Yes | No |
| **Neurological disorders** | | | | | | |
| Three-dimensional convolutional neural network algorithm to classify neurological disorders as critical or non-critical for triage[41] | Imaging data (CT brain scan) | Prospective, randomised clinical trial showed sensitivity comparable to clinicians but much lower specificity. However, the study only involved one hospital site and requires external validation | No | Yes | Yes | No |
| **Septic shock** | | | | | | |
| TREWScore for predicting risk for septic shock in ICU and non-ICU patients[42,43] | Risk prediction based on electronic health record clinical and lab data | Uses routine clinical findings and lab results from electronic health records to derive risk scores that predict septic shock before onset, with a sensitivity >80% in both ICU and non-ICU patients. Specificity was 67% in ICU and 90% in non-ICU patients, greater than that for a conventional screening protocol for septic shock | No | No | No | No |
| **Oncology** | | | | | | |
| PowerLook Density Assessment 3.4 (iCAD, Nashua, NH) for automated breast density assessment, an important factor in breast cancer diagnosis[44] | Imaging data (mammogram images) | The algorithm produced similar assessments as radiologists, but with higher reproducibility. Limitations are that the algorithm was not validated for tomosynthesis (three-dimensional mammography) and that only white and Asian women were studied | Yes | No | Yes | Yes |
| Cloud-based deep neural network algorithms (Oncology Lung AI and Oncology Liver AI, Arterys, San Francisco, CA) for detection and segmentation of lung and liver tumours[45] | Imaging data (CT) | Detection and segmentation performance shown to be comparable with that of expert radiologists; tools were designed to be used in a workflow that involves clinician review and possible modification of results before accepting them. Lack of peer-reviewed studies makes it difficult to fully evaluate the algorithm | Yes | Yes | Yes | Yes |
| Deep neural network algorithm for detecting malignant lung nodules[46] | Imaging data (radiographs) | Performed better than most of the 18 radiologists it was compared with in a retrospective study, and improved performance of radiologists when used in conjunction with them. Drawbacks include the retrospective nature of the evaluation and the lack of benign nodules in the training set, thus the algorithm cannot optimally differentiate benign from malignant nodules | Yes | Yes | Yes | No |

(Table 1 continues on next page)

| | Type of data used or clinical application | Comments | External validation | Validated with humans | Validated against humans | US Food & Drug Administration approval |
|---|---|---|---|---|---|---|
| (Continued from previous page) | | | | | | |
| Deep neural network algorithm for grading prostate cancer tumours[47] | Imaging data (ultrasound of biopsy cores) | Uses learned features from temporal enhanced ultrasound images, followed by distribution learning, to classify prostate tumours into three categories: aggressive prostate cancer, clinically less significant prostate cancer, and non-cancerous prostate tissue. The tool could be successfully trained to produce multilabel readouts (non-trivial) but has only been evaluated for biopsies that were suspicious by MRI | No | No | Yes | No |
| Deep neural network that produces glioma survival predictions[48] | Histological imaging and genomic marker data | Considers both histological and genomic data in a single survival prediction framework, and did as well as conventional histological grading and molecular subtyping methods. A limitation is that the tool was trained on a relatively small portion (regions of interest) of each histological slide, and experts were needed to select these regions of interest | No | No | Yes | No |
| Watson for Oncology (IBM, Armonk, NY) is an algorithm that produces recommendations for cancer treatments[49] | Clinical data from patients with cancer | According to IBM, Watson for Oncology has been trained to provide safe treatment recommendations for 13 cancers. However, it produced inaccurate predictions that contradicted treatment guidelines and doctors' medical judgment[38] | Yes | No | No | No |
| CURATE.AI provides predictions for optimal drug doses for patients with prostate cancer[50] | Medication dosage and treatment response data from electronic health records | Guided dynamic dosing of a novel combination therapy for prostate cancer, improving tolerance and efficacy of the therapy. However, its application in guiding continuous dosing has been tested in just one patient and will benefit from larger studies | No | No | No | No |

The machine learning algorithms described in the table were selected from algorithms developed at National University Health System and those reviewed in the *PLoS Medicine* Machine Learning in Health and Biomedicine Special Issue (March 1, 2018) and in Topol.[11] The algorithms include health-care tools in oncology and other applications that use different types of data and are at various stages of clinical evaluation and development. ICU=intensive care unit

***Table 1:*** **Machine learning algorithms for different clinical applications**

| | Intervention | Category | Status | Conditions | Country | Clinicaltrials.gov registry number |
|---|---|---|---|---|---|---|
| Artificial Intelligence *vs* Physicians for Breast Cancer Patients' Information | Chatbot to provide information to patients with breast cancer *vs* multidisciplinary medical committee | Chatbot | Not yet recruiting | Breast cancer | France | NCT03556813 |
| An Enhanced Artificial Intelligence Breast MRI Interpretation System | Machine learning system to interpret breast MRI | Machine vision | Not yet recruiting | Breast cancer | UK | NCT03829423 |
| Breast Ultrasound Image Reviewed with Assistance of Deep Learning Algorithms | Machine learning tool classifying breast lesions using breast ultrasound images | Machine vision | Recruiting | Breast cancer | USA | NCT03706534 |
| Development of Artificial Intelligence System for Detection and Diagnosis of Breast Lesion Using Mammography | Development of a machine learning tool to classify breast lesions using breast mammograms | Machine vision | Recruiting | Breast cancer | China | NCT03708978 |
| Adenoma Detection Rate Using AI System in China | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Not yet recruiting | Colonoscopy for colorectal polyp | China | NCT03840590 |
| Artificial Intelligence Identifying Polyps in Real-World Colonoscopy | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Completed | Colonoscopy for colorectal polyp | China | NCT03761771 |
| Automatic Classification of Colorectal Polyps Using Probe-Based Endomicroscopy with Artificial Intelligence | Machine learning tool classifying colorectal polyps using probe-based confocal laser endomicroscopy images | Machine vision | Recruiting | Colonoscopy for colorectal polyp | China | NCT03787784 |
| Computer Assisted Detection and Selection of Serrated Adenomas and Neoplastic Polyps—a New Clinical DRAft | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Not yet recruiting | Colonoscopy for colorectal polyp | Germany | NCT03601065 |
| Computer-Aided Detection for Colonoscopy | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Not yet recruiting | Colonoscopy for colorectal polyp | Taiwan | NCT03842059 |
| Deep-Learning for Automatic Polyp Detection During Colonoscopy | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Recruiting | Colonoscopy for colorectal polyp | USA | NCT03637712 |
| Diagnostic Performance of a Convolutional Neural Network for Diminutive Colorectal Polyp Recognition | Development of a machine learning tool to classify colorectal polyps using endoscopy images | Machine vision | Active, not recruiting | Colonoscopy for colorectal polyp | Netherlands | NCT03822390 |

(Table 2 continues on next page)

| | Intervention | Category | Status | Conditions | Country | Clinicaltrials.gov registry number |
|---|---|---|---|---|---|---|
| (Continued from previous page) | | | | | | |
| In Vivo Computer-aided Prediction of Polyp Histology on White Light Colonoscopy | Machine learning tool classifying colorectal polyps using endoscopy images | Machine vision | Not yet recruiting | Colonoscopy for colorectal polyp | Spain | NCT03775811 |
| AI-EMERGE: Development and Validation of a Multi-analyte, Blood-based Colorectal Cancer Screening Test | Machine learning used to identify biomarkers of early colorectal cancer in cell-free blood assays | Biomarker discovery | Recruiting | Colorectal cancer | USA | NCT03688906 |
| Automatic Real-time Diagnosis of Gastric Mucosal Disease Using pCLE With Artificial Intelligence | Machine learning tool classifying gastric mucosal disease using probe-based confocal laser endomicroscopy images | Machine vision | Recruiting | Gastric cancer | China | NCT03784209 |
| Clinical Concordance Study Between Watson for Oncology and Clinician Practice | Machine learning used to recommend treatments for patients with cancer *vs* physicians' recommendations | Clinical matching | Recruiting | General cancer | China | NCT03400514 |
| SYNERGY-AI: Artificial Intelligence Based Precision Oncology Clinical Trial Matching and Registry | Clinical trial matching using machine learning algorithm | Clinical matching | Recruiting | General cancer | USA | NCT03452774 |
| Evaluation of Lung Nodule Detection with Artificial Intelligence Assisted Computed Tomography in North China | Development of a machine learning tool to detect lung nodules in low-dose lung CT screening | Machine vision | Not yet recruiting | Lung cancer | China | NCT03487952 |
| IDEAL: Artificial Intelligence and Big Data for Early Lung Cancer Diagnosis Study | Computer-aided prediction tool to diagnose lung nodules on lung CT | Machine vision | Recruiting | Lung cancer | UK | NCT03753724 |
| Lung Nodule Characterization by Artificial Intelligence Techniques | Machine learning tool to detect lung nodules on lung CT | Machine vision | Not yet recruiting | Lung cancer | France | NCT03843164 |
| Application of Genomic Techniques and Image Processing Using Artificial Intelligence to Obtain a Predictor Model Risk of Melanoma | Machine learning algorithm to predict risk of melanoma using genomic and imaging data | Machine vision | Unknown status | Melanoma | Spain | NCT02511119 |
| Artificial Intelligence-assisted Evaluation of Pigmented Skin Lesions | Machine learning tool to classify dermoscopic images to classify dysplastic nevi, spitz nevi, and malignant melanoma | Machine vision | Recruiting | Melanoma | Israel | NCT03362138 |
| Elucid Labs AIDA—Labelled Image Acquisition Protocol | Machine learning algorithm to evaluate melanoma as well as basal and squamous cell carcinoma | Machine vision | Not yet recruiting | Melanoma | Canada | NCT03621462 |
| Artificial Intelligence for Early Diagnosis of Esophageal Squamous Cell Carcinoma | Endoscopic image-analysing machine learning system to detect early oesophageal cancer | Machine vision | Recruiting | Oesophageal cancer | China | NCT03759756 |
| Project Survival-Prospective Biomarker Discovery | Machine learning used to identify biomarkers of early pancreatic cancer from tissues and fluid samples | Biomarker discovery | Recruiting | Pancreatic cancer | USA | NCT02781012 |
| Developing an Imaging-Based Tool to Identify Areas for Prostate Cancer Biopsy | Machine learning tool to identify areas for biopsy in prostate cancer using prostate MRI images | Machine vision | Recruiting | Prostate cancer | USA | NCT03585660 |
| Profiling of Radiological Factors in Treatment and Outcomes in Prostate Cancer | Development of a machine learning tool using multiparametric MRI of prostate glands, correlated with histology | Machine vision | Recruiting | Prostate cancer | USA | NCT03354416 |
| PSMA-PET: Deep Radiomic Biomarkers of Progression and Response Prediction in Prostate Cancer | Development of machine learning tool to predict survival of patients with prostate cancer using prostate-specific membrane antigen-PET imaging | Machine vision | Recruiting | Prostate cancer | Canada | NCT03594760 |

The information in this table was obtained by searching ClinicalTrials.gov for all countries and time periods using the terms: "artificial Intelligence", "AI", "algorithms", "convoluted neural networks", "deep learning", "machine learning", and "neural networks". Only oncology-related trials were selected and screened for inclusion, on the basis of the use of machine learning methods as interventions.

***Table 2:*** **Clinical trials evaluating machine learning algorithms for oncological applications**

## Examples of machine learning platforms

On the long road towards deployable clinical machine learning tools, every aspect of the build must be addressed and integrated to ensure the safety and efficacy of the tools used in clinical practice. Ideally, these aspects should exist in a single platform for ease of access, governance, and oversight.

Several such platforms have been attempted, including the Philips (Amsterdam, Netherlands) HealthSuite Insights platform, which brings together machine learning tools and analytic capabilities in diagnostic imaging, patient monitoring, genomics, and oncology to facilitate their deployment for clinical and research applications. Health-care professionals can access different types of patient data, curate the data, and apply it to personalised diagnosis and treatment.[53]

Another platform, GE Healthcare's (Chicago, IL) Edison, connects data from millions of imaging devices, such as MRI and CT machines, facilitating faster and more accurate diagnoses based on imaging data. One of the major uses of Edison is in updating existing equipment in hospitals by plugging them into applications and algorithms with data analytics capabilities.[54]

Hitachi's (Tokyo, Japan) Lumada platform uses machine learning data analytics as well as Internet of Things (remote monitoring of equipment) and control technologies for various applications, including health care. As a use case, the Lumada platform helped to improve prediction of the failure of MRI machines, enabling pre-emptive maintenance and improved machine performance.[55]

The DISCOVERY AI platform was developed at the National University Health System (Singapore) as a production system that houses modular machine learning tools (figure 2). The platform uses daily data from the electronic health record system to make predictions about patients. Simultaneously, it also acts as a research sandbox by linking and aggregating multiple clinical and research databases to facilitate the joint development of machine learning tools.

A key feature of the platform is its unique master governance structure, which ensures equitable permissioned data access, centralised anonymisation, and differential data linkage. The same master access control also acts as the central trusted third party to administer these features. Oversight of data access and sharing rests with the custodian of a particular database and a committee governing data access and sharing, as well as the linkage of a particular database to other databases. Research administration of the AI tools, including institutional review board processes, data deposit agreements, and research collaborative agreements, also falls under the same governance structure.

Besides meeting enterprise security requirements, DISCOVERY AI incorporates centralised anonymisation and data handling measures that are in accordance with the Singapore Personal Data Protection Act 2012,[56] Human Biomedical Research Act 2015,[57] and Human Biomedical Research Regulations 2017.[58] In keeping with Personal Data Protection Act guidelines, all data on board DISCOVERY AI are anonymised by removing identifiers, such as name, address and postal code, and identification number, and can be used for analysis and training of AI systems. Additionally, the platform features proprietary security features, such as data obfuscation and ledger-based access logs.

The machine training aspects of developing a machine learning health-care tool, such as automated data pre-processing, clinical data curation, the machine learning framework, and deep learning best practices, are integral to the development process of machine learning tools within the platform. The computing architecture of the platform is able to host multiple modular machine learning systems, all using data from the same electronic health record source (figure 2).

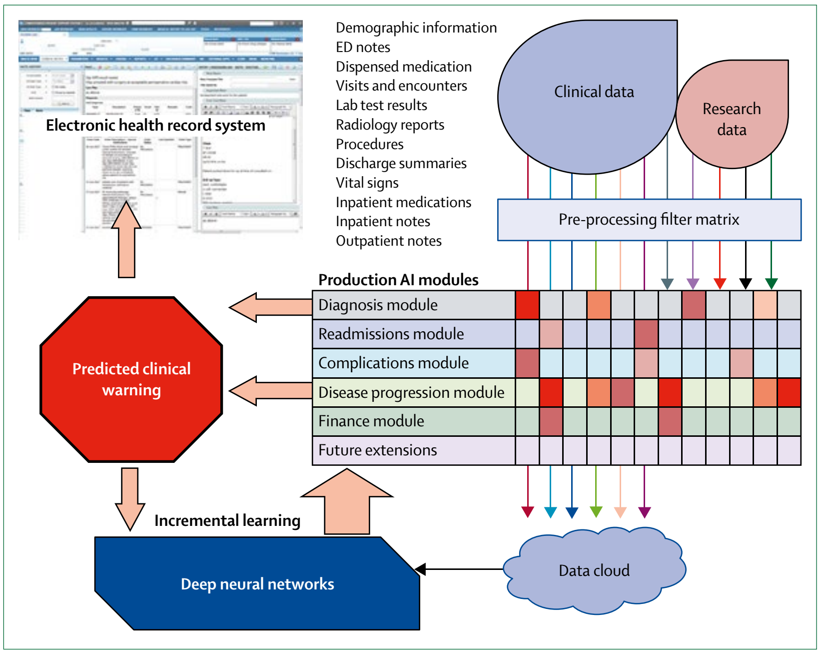


*Figure 2:* **DISCOVERY AI platform processes**
DISCOVERY AI houses various modular AI tools that process clinical and research data to make predictions, which can then be sent as clinical alerts to the electronic health record system. AI=artificial intelligence.

This juxtaposition of machine learning modules in one system allows machine learning tool designers to aggregate the outputs of various machine learning algorithms into fewer alerts, thus reducing alert fatigue for doctors.

The integration of machine learning-driven alerts with the use of the electronic health record system allows the capture of doctors' responses to machine alerts and completes the closed data feedback loop. This avoids the need for doctors' responses to be manually captured and provides the data required for incremental training of machine learning tools. With this system in place, clinical validation trials can be done in production systems and the claims of the component machine learning tools can be tested, with the results forming the basis for future applications for regulatory approvals. The integrated system can also gather valuable feedback from doctors, simultaneously educating and canvassing their endorsement of these new tools.

## Examples of machine learning applications in oncology care delivery

CURATE.AI is a machine learning application developed at the National University of Singapore and the University of California Los Angeles that helps doctors to optimise drug dosages, especially for combination therapy. It does so by using data about a patient's drug doses and responses over time to continuously predict optimal drug doses for that patient. CURATE.AI has aided the treatment of a patient with metastatic castration-resistant prostate cancer using a novel combination of enzalutamide and the investigational drug ZEN-3694.[50] An individualised profile was created for the patient of the relationship between drug dose and efficacy (reduction in prostate-specific antigen levels) over the initial 6 months of treatment, when his doctors adjusted the dosages of the drugs to manage his personal adverse event profile. This profile showed a strong correlation between drug dose and efficacy ($R^2=0 \cdot 96$) and the high actionability of the recommendations ($R^2=0 \cdot 98$).

On the basis of the patient's dose-efficacy profile, CURATE.AI recommended an initial dose for ZEN-3694 that was half of the initial dose determined by the patient's doctors before the AI tool was used. CURATE.AI also produced recommendations for optimal doses for both drugs throughout the treatment course that were based on the patient's profile, which was regularly updated to include changes in the patient's condition and prostate-specific antigen levels. The patient's clinical team approved each dosage adjustment before it was made and, in some instances, made decisions on the basis of their clinical knowledge (eg, maintaining the same dose for one of the drugs at the start of the study). Aided by CURATE.AI-guided dynamic dosing, the novel combination therapy was well tolerated and effective, reducing the patient's prostate-specific antigen concentrations and halting disease progression.

Another study[59] of eight liver transplant patients whose immunosuppressant drug dosing was guided by CURATE.AI had less variability in drug trough levels compared with patients who received standard-of-care dosing. More clinical trials to evaluate this tool in other applications, including more oncological applications, are ongoing (NCT03759093).

Another machine learning tool developed at National University Health System, Augurium, performs natural language processing of doctors' notes to help improve the accuracy of appendicitis diagnoses. This scalable tool was trained on data from 200 000 cases presenting with abdominal pain.

These complex machine learning tools offer great promise for improving health-care delivery through the provision of personalised recommendations to doctors, based on insights derived from large, longitudinal datasets. However, many such tools have not been subjected to a rigorous peer-review process. Because they are trained with population-specific datasets, they might not be immediately generalisable to all other populations unless they are subsequently trained with data from these other populations or if newer tools arise.

## Discussion

By exploiting capabilities such as deep neural networks, machine learning presents a powerful tool for analysing large amounts of complex health-care data to improve the efficiency and cost-effectiveness of health-care delivery. When used to augment the capabilities of doctors, machine learning can perform routine, systematic tasks with high consistency, freeing up doctors' time so that they can address clinical problems that are more complex or that require substantial human interaction.

However, machine learning tools have several limitations or requirements that need to be addressed before they can be effectively deployed in health care. The algorithms require data pre-processing, training on datasets (which might be large), and iterative refinement with respect to the actual clinical problem. Adopting machine learning algorithms in clinical practice also raises several key ethical concerns. These concerns include liability in cases of medical error, doctors' understanding of how machine learning tools produce predictions, patients' understanding and control of how machine learning tools are used in their care, and issues of privacy, security, and control of patient data.

The use of machine learning algorithms to augment rather than replace the doctor helps to increase clarity for medical liability issues. In instances of medical error, the liability still rests with the doctor as the clinical tasks either involve or are supervised by a doctor, although legal experts are still debating whether some liability should rest with the developer of the machine learning tool as well.[60] When doctors override machine learning-based predictions, the law requires doctors to show that their actions are reasonable and consistent with those of a typical member of their profession.[61] It could be argued that machine learning tools should be beyond such standard practice obligations, but this issue could change as these tools become more prevalent in the clinical decision-making process.

Although doctors might not need to know the detailed mathematical calculations used in a machine learning algorithm, they could be educated about the types of data used in making the predictions and the relative weights assigned to each type of data (if the machine learning tool is designed to be explainable). Similar to a clinical test result, doctors could consider features such as sensitivity and specificity for predicting a particular disease risk or treatment response. Some countries are also recognising the patient's right to understanding machine learning tools when they are used in their health care.[62]

Machine learning tools often use sensitive personal data to make stratified health-care recommendations. This use of personal data raises considerations of ensuring privacy and security of the data,[62] transparent communication with the public about uses of their data, and protecting against the use of data or machine learning algorithms in discriminatory practices.

To address some of these concerns, platforms that enable the building of multiple machine learning tools and feature linkage to a patient's electronic health record

## Search strategy and selection criteria

References were first identified by searching articles in PubMed from Jan 1, 1995, until Oct 31, 2018, using the search terms "artificial intelligence" OR "AI" OR "AI-assisted" OR "machine learning" OR "deep learning" OR "deep neural networks" OR "algorithms" AND "big data" AND ("healthcare" OR "medicine"). This search produced 643 articles, of which 22 were relevant to machine learning algorithms in oncology, that were selected for inclusion in the final reference list because of their relevance to the topic or ability to provide new insights. Additional references about specific topics were also included.

and data governance systems offer a promising model for the future. Numerous machine learning tools can be custom developed for disease-specific indications, such as predicting breast cancer recurrences and hosting pharmacogenomics workflows for chemotherapy drugs. Machine learning tools can also be built for system-level interventions, including improving patient selection and recruitment for clinical trials, reducing patient readmission, and automated follow-up of patients for surveillance of complications. Similar to other clinical tools, these machine learning platforms will need to be evaluated as part of the clinical workflow in real-world health-care settings.

### Declaration of interests

We declare no competing interests.

### References


1. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 1959; **3:** 210–29.
2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.
3. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 2014; **33:** 1123–31.
4. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. *Comput Methods Programs Biomed* 2018; **161:** 1–13.
5. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA* 2016; **316:** 2353–54.
6. Kantarjian H, Yu PP. Artificial intelligence, big data, and cancer. *JAMA Oncol* 2015; **1:** 573–74.
7. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. *Nat Biotechnol* 2018; **36:** 829–38.
8. Mazzanti M, Shirka E, Gjergo H, Hasimi E. Imaging, health record, and artificial intelligence: hype or hope? *Curr Cardiol Rep* 2018; **20:** 48.
9. Personal Data Protection Commission Singapore. A Proposed model artificial intelligence governance framework. January 2019. Personal Data Protection Commission Singapore, 2019. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf (accessed April 12, 2019).
10. Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. *Front Neurol* 2017; **8:** 489.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25:** 44–56.
12. Gelhaus P. Robot decisions: on the importance of virtuous judgment in clinical decision making. *J Eval Clin Pract* 2011; **17:** 883–87.
13. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. *Artif Intell Med* 2014; **62:** 1–10.
14. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342:** 1887–92.
15. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Inform* 2010; **79:** 736–71.
16. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. *N Engl J Med* 2017; **377:** 1585–92.
17. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542:** 115–18.
18. Kowatsch T, Nissen M, Chen-Hsuan IS, et al. Text-based healthcare chatbots supporting patient and health professional teams: preliminary results of a randomized controlled trial on childhood obesity. Persuasive Embodied Agents for Behavior Change (PEACH2017) Workshop, co-located with the 17th International Conference on Intelligent Virtual Agents (IVA 2017); Stockholm, Sweden; Aug 27–30, 2017.
19. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018; **25:** 1248–58.
20. Zheng KP, Gao J, Ngiam KY, Ooi BC, Yip WLJ. Resolving the bias in electronic medical records. 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Halifax, Nova Scotia, Canada; Aug 13–17, 2017.
21. Gawehn E, Hiss JA, Brown JB, Schneider G. Advancing drug discovery via GPU-based deep learning. *Expert Opin Drug Discov* 2018; **13:** 579–82.
22. Obeid NM, Atkinson IC, Thulborn KR, Hwu W-MW. GPU-accelerated gridding for rapid reconstruction of non-cartesian MRI. 19th Annual International Society for Magnetic Resonance in Medicine (ISMRM) Scientific Meeting and Exhibition 2011; Montreal, QC, Canada; May 7–13, 2011.
23. Wang H, Peng H, Chang Y, Liang D. A survey of GPU-based acceleration techniques in MRI reconstructions. *Quant Imaging Med Surg* 2018; **8:** 196–208.
24. van Grinsven M, van Ginneken B, Hoyng C, Theelen T, Sanchez C. Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans Med Imag* 2016; **35:** 1273–84.
25. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. *N Engl J Med* 2018; **379:** 1452–62.
26. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; **32:** D267–70.
27. Marcus G. Deep learning: a critical appraisal. *arXiv* 2018; 1801.00631.
28. Kuo C-CJ. Understanding convolutional neural networks with a mathematical model. *arXiv* 2016; 1609.04112.
29. Deo RC. Machine learning in medicine. *Circulation* 2015; **132:** 1920–30.
30. Graves A, Mohamed A-R, Hinton G. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada; May 26–31, 2013.
31. Greenspan H, van Ginneken B, Summers RM. Deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imag* 2016; **35:** 1153–59.
32. US Food & Drug Administration. Changes to existing medical software policies resulting from section 3060 of the 21st Century Cures Act: draft guidance for industry and Food and Drug Administration staff. Silver Spring, MD, USA: US Food & Drug Administration, 2017.
33. US Food & Drug Administration. Software as a medical device (SAMD): clinical evaluation. Guidance for industry and Food and Drug Administration staff. Silver Spring, MD, USA: US Food & Drug Administration, 2017.
34. Gepperth A, Hammer B. Incremental learning algorithms and applications. European Symposium on Artificial Neural Networks (ESANN) 2016; Bruges, Belgium; April 27–29, 2016.
35. Silver DL. Machine lifelong learning: challenges and benefits for artificial general intelligence. Artificial General Intelligence (AGI) 2011; Mountain View, CA, USA; Aug 3–6, 2011.
36. Abràmoff MD, Lavin PT, Birch M, Shahm N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digit Med* 2018; **1:** 39.

37 Lin H, Long E, Ding X, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS Med* 2018; **15:** e1002674.

38 Marr B. First FDA approval for clinical cloud-based deep learning in healthcare. Forbes, Jan 20, 2017. https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#2e0b1a44161c (accessed Feb 19, 2019).

39 Voelker R. Diagnosing fractures with AI. *JAMA* 2018; **320:** 23.

40 Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018; **15:** e1002686.

41 Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med* 2018; **24:** 1337–41.

42 Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; **7:** 299ra122.

43 Henry K, Wongvibulsin S, Zhan A, Saria S, Hager D. Can septic shock be identified early? Evaluating performance of a targeted real-time early warning score (TREWScore) for septic shock in a community hospital: global and subpopulation performance *Am J Resp Crit Care Med* 2017; **195:** A7016 (abstr).

44 Kerlikowske K, Scott CG, Mahmoudzadeh AP, et al. Automated and clinical breast imaging reporting and data system density measures predict risk for screen-detected and interval cancers: a case-control study. *Ann Intern Med* 2018; **168:** 757–65.

45 Arterys FDA clearance for Liver AI and Lung AI lesion spotting software. *Medgadget*, Feb 19, 2018. https://www.medgadget.com/2018/02/arterys-fda-clearance-liver-ai-lung-ai-lesion-spotting-software.html (accessed Feb 19, 2019).

46 Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019; **290:** 218–28.

47 Azizi S, Bayat S, Yan P, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. *Int J Comput Assist Radiol Surg* 2017; **12:** 1293–305.

48 Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci USA* 2018; **115:** E2970–79.

49 Ross C, Swetlitz I. IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. STAT, 2018. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/ (accessed March 19, 2019).

50 Pantuck AJ, Lee D-K, Kee T, et al. Modulating BET bromodomain inhibitor ZEN-3694 and enzalutamide combination dosing in a metastatic prostate cancer patient using CURATE.AI, an artificial intelligence platform. *Adv Therap* 2018; **1:** 1800104.

51 US Food & Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Silver Spring, MD, USA: US Food & Drug Administration, 2018. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm (accessed March 19, 2019).

52 Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017; **2:** 230–43.

53 Philips. Philips launches AI platform for healthcare. Philips, 2018. https://www.philips.com/a-w/about/news/archive/standard/news/press/2018/20180301-philips-launches-ai-platform-for-healthcare.html (accessed March 19, 2019).

54 GE Healthcare. One of the largest AI platforms in healthcare is one you've never heard of, until now. The Pulse, 2018. http://newsroom.gehealthcare.com/new-apps-smart-devices-launch-healthcare-edison-ai-platform/ (accessed March 19, 2019).

55 Hitachi. Predictive maintenance of medical devices based on years of experience and advanced analytics. Hitaci, 2019. http://social-innovation.hitachi/en/case_studies/mri_predictive_maintenance/?__CAMCID=lknjlhToJY-387&__CAMSID=cUeDHEgFEGyU-74&__CAMVID=EfODhEgFEGYU&_c_d=1&_ct=1548576884716 (accessed Jan 20, 2019).

56 Government of the Republic of Singapore. Personal Data Protection Act 2012. Singapore: Government of the Republic of Singapore, 2012.

57 Government of the Republic of Singapore. Human Biomedical Research Act 2015. Singapore: Government of the Republic of Singapore, 2015.

58 Government of the Republic of Singapore. Human Biomedical Research Regulations 2017. Singapore: Government of the Republic of Singapore, 2017.

59 Zarrinpar A, Lee DK, Silva A, et al. Individualizing liver transplant immunosuppression using a phenotypic personalized medicine platform. *Sci Transl Med* 2016; **8:** 333ra49.

60 Loh E. Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. *BMJ Leader* 2018; **2:** 59–63.

61 McNair J, Ottley, R. Negligence: who is the umpire. *J Med Defence Union* 1996; **10:** 18–20.

62 Daniel GW, Silcox CE, Sharma I, Wright MB. Current state and near-term priorities for AI-enabled diagnostic support software in health care. Margolis Center for Health Policy, 2019. https://healthpolicy.duke.edu/sites/default/files/atoms/files/dukemargolisaienableddxss.pdf (accessed March 19, 2019).