

How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices

David Lyell , Enrico Coiera, Jessica Chen, Parina Shah, Farah Magrabi

To cite: Lyell D, Coiera E, Chen J, *et al*. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. *BMJ Health Care Inform* 2021;**28**:e100301. doi:10.1136/bmjhci-2020-100301

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2020-100301>).

Received 07 December 2020
Revised 10 February 2021
Accepted 02 March 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia

Correspondence to

Dr David Lyell;
david.lyell@mq.edu.au

ABSTRACT

Objective To examine how and to what extent medical devices using machine learning (ML) support clinician decision making.

Methods We searched for medical devices that were (1) approved by the US Food and Drug Administration (FDA) up till February 2020; (2) intended for use by clinicians; (3) in clinical tasks or decisions and (4) used ML. Descriptive information about the clinical task, device task, device input and output, and ML method were extracted. The stage of human information processing automated by ML-based devices and level of autonomy were assessed.

Results Of 137 candidates, 59 FDA approvals for 49 unique devices were included. Most approvals (n=51) were since 2018. Devices commonly assisted with diagnostic (n=35) and triage (n=10) tasks. Twenty-three devices were assistive, providing decision support but left clinicians to make important decisions including diagnosis. Twelve automated the provision of information (autonomous information), such as quantification of heart ejection fraction, while 14 automatically provided task decisions like triaging the reading of scans according to suspected findings of stroke (autonomous decisions). Stages of human information processing most automated by devices were information analysis, (n=14) providing information as an input into clinician decision making, and decision selection (n=29), where devices provide a decision.

Conclusion Leveraging the benefits of ML algorithms to support clinicians while mitigating risks, requires a solid relationship between clinician and ML-based devices. Such relationships must be carefully designed, considering how algorithms are embedded in devices, the tasks supported, information provided and clinicians' interactions with them.

INTRODUCTION

Artificial intelligence (AI), technologies undertake recognition, reasoning or learning tasks typically associated with human intelligence,¹ such as detecting disease in an image, diagnosis and recommending treatments, have the potential to improve healthcare delivery and patient outcomes.² Machine learning (ML) refers more specifically to AI methods that can learn from data.³ The

Summary

What is already known?

- Machine learning (ML)-based clinical decision support (CDS) operates within a human–technology system.
- Clinician interaction with CDS influences how they make decisions affecting care delivery and patient safety.
- Little is known about how emerging ML-based CDS supports clinician decision making.

What does this paper add?

- ML-based CDS approved by the FDA typically provide clinicians with decisions or information to support their decision making.
- Most demonstrate limited autonomy, requiring clinicians to confirm information provided by CDS and to be responsible for decisions.
- We demonstrate methods to examine how ML-based CDS are used by clinicians in the real world.

current resurgence in ML is largely driven by developments in deep learning methods, which are based on neural networks. Despite the expanding research literature, relatively little is known about how ML algorithms are embedded in working clinical decision support (CDS).

CDS that diagnoses or treats human disease automate clinical tasks otherwise done by clinicians.⁴ Importantly, CDS operates within a human–technology system,⁵ and clinicians can elect to ignore CDS advice and perform those tasks manually. Clinician interaction with ML-based CDS influences how they work and make decisions which in turn affects care quality and patient safety.

Alongside intended benefits, ML poses new risks that require specific attention. A fundamental challenge is that ML-based CDS may not generalise well beyond the data on which they are trained. Even for restricted tasks like image interpretation, ML algorithms

can make erroneous diagnoses because of differences in the training and real-world populations, including new ‘edge’ cases, as well as differences in image capture workflows.⁶ Therefore, clinicians will need to use ML-based CDS within the bounds of their design, monitor performance and intervene when it fails. Clinician interaction with CDS is thus a critical point where the limitations of ML algorithms are either mitigated or translated into harmful patient safety events.^{7–9}

One way to study the interaction between clinicians and ML-based CDS is to consider medical devices. In the USA, software, including CDS that is intended to diagnose, cure, mitigate, treat or prevent disease in humans, are considered medical devices¹⁰ and subject to regulation. Increasing numbers of devices that embody ML algorithms are being approved by the US Food and Drug Administration (FDA).^{11 12} Approval requires compliance with standards, as well as evaluation of device safety and efficacy.¹³ Regulators provide public access to approvals and selected documentation. Therefore, medical devices provide a useful sample for studying how ML algorithms are embedded into CDS for clinical use and how manufacturers intend clinicians to interact with them.

Research has predominantly focused on the development and validation of ML algorithms, and evaluation of their performance,^{11 14–16} with little focus on how ML is integrated into clinical practice and the human factors related to their use.¹⁷ In a recent systematic review of ML in clinical medicine, only 2% of studies were prospective, most were retrospective providing ‘proof of concept’ for how ML might impact patient care, without comparison to standard care.¹⁸

While one recent study has described the general characteristics of 64 ML-based medical devices approved by the FDA,¹² no previous study has examined how ML algorithms are embedded to support clinician decision making. Our analysis of ML medical devices thus seeks to bridge the gap between ML algorithms and how they are used in clinical practice.

Human information processing

In assessing human–machine interaction, it is useful to consider how clinicians process information and make decisions, and which stages of that process are automated by ML devices. Automaton is the machine performance of functions otherwise be done by humans.⁴ Human information processing has been broken down into four distinct stages: (1) Sensing information in the environment, (2) Perceiving or interpreting what the information means, (3) Deciding the appropriate response and (4) Acting on decisions (figure 1).¹⁹ For example, the diagnosis of pneumonia requires clinicians to sense information relevant to the provisional and differential diagnoses of the patient’s condition from their medical history, physical examination and diagnostic tests. Information then needs to be interpreted: do chest X-rays show evidence of inflammation? These analyses inform

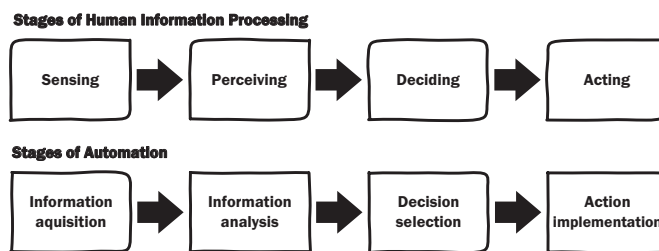


Figure 1 Stages of human information processing (top) and their automation (bottom).¹⁹

decisions about diagnosis and treatment, which are then enacted by ordering or referring for treatment.

ML devices can automate any or all stages of human information processing: (1) Acquiring information, (2) Analysing information, (3) Decision selection from available alternatives and (4) Implementation of the selected decision (figure 1).¹⁹ Later stages represent higher levels of automation. For instance, an ML device assisting the diagnosis of cardiac arrhythmias that report quantitative measurements from ECGs, automates information analysis, whereas devices that indicate the presence or absence of atrial fibrillation automate decision selection. Identifying the stage of human information processing automated provides a useful framework for evaluating how ML devices change clinicians’ work, especially the division of labour between clinicians and ML devices.

Accordingly, we examined FDA-approved ML devices to understand:

- ▶ Which ML devices have been approved for clinical practice, their intended use, the diseases they diagnose, treat or prevent, and how manufacturers intend for clinicians to interact with them?
- ▶ How ML devices might change clinician decision making by exploring the stage of human information processing automated.
- ▶ The extent to which ML devices function autonomously and how that impacts clinician–ML device interaction.

METHOD

We examined FDA-approved medical devices that use ML (online supplemental appendix A). Unable to directly search FDA databases for ML devices, we used an internet search to identify candidate devices that were:

1. FDA-approved medical devices.
2. Intended for use by clinicians.
3. Intended to support clinical tasks/decisions.
4. Using ML.

The search identified 137 candidate devices for which 130 FDA approvals were retrieved. Of these, 59 approvals met the inclusion criteria covering 49 unique ML devices (figure 2).

Data extraction and analysis

For each included approval, we extracted the approval details (date, pathway, device risk class). For each unique device, we then extracted type (software as a medical

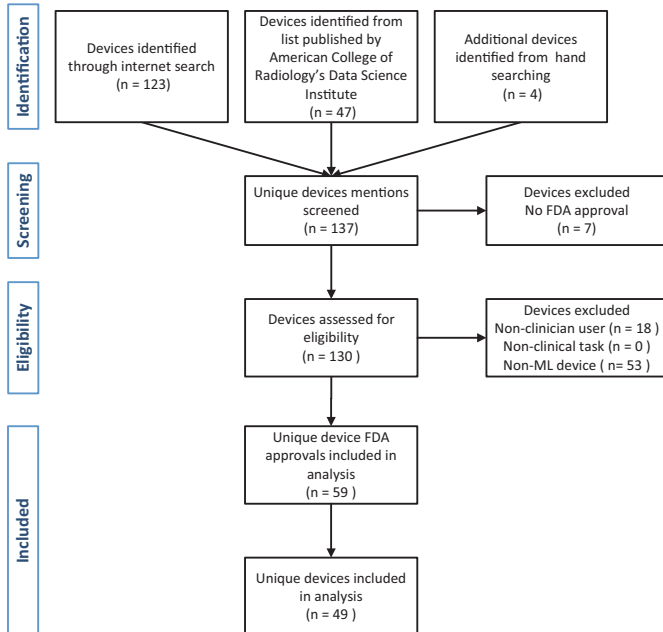


Figure 2 Process to search for and identify FDA Approved ML devices. FDA, Food and Drug Administration; ML, machine learning.

device or integrated into hardware); characteristics (indicated disease, clinical task, device task, input, output); and ML method used as described in the approval. Clinical task, device input and output were identified from device indications and descriptions, and grouped according to natural categories emerging from the sample. The device task was summarised from the indications and device description in FDA approvals.

Stage of human information processing automated by ML devices

The device task was examined using the stages of automation of human information processing framework.¹⁹ We classified the highest stage of human information processing (figure 1) automated by ML devices according to the following criteria (from lowest to highest):

1. Information acquisition: Device automates data acquisition and presentation for interpretation by clinicians. Data are preserved in raw form, but the device may aid presentation by sorting, or enhancing data.
2. Information analysis: Device automates data interpretation, producing new information from raw data. Importantly, interpretation contributes new information that supports decision making, without providing the decision. For example, the quantification of QRS duration from electrocardiograms provides new information from ECG tracings that may inform diagnosis without being a diagnosis.
3. Decision selection: Device automates decision making, providing an outcome for the clinical task. For example, prompting and thereby drawing attention to malignant lesions on screening mammograms, indicates a device decision about the presence of breast cancer.
4. Action implementation: Device automates implementation of the selected decision where action is required.

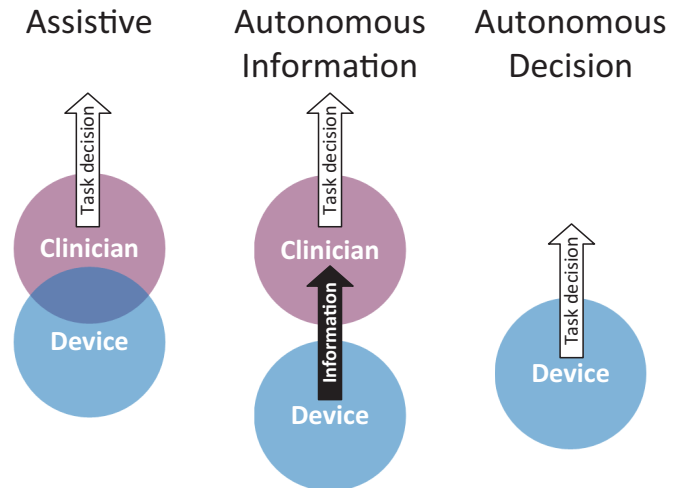


Figure 3 Level of autonomy showing the relationship between clinician and device.

For example, an implantable cardioverter-defibrillator, having decided defibrillation is required, acts by automatically delivering treatment.

ML device autonomy

To understand the level of device autonomy, we examined the description and indications for use to determine the extent to which the device performs automated tasks independent of clinicians.¹⁹ For example, a device automating decision selection that requires clinician approval is less autonomous than similar devices that do not require approval. The approach is similar to existing levels of autonomy for specific tasks, such as driving automation,²⁰ and computer-based automation,²¹ which identify what user and automation are responsible for in relation to a defined task. Taking these models as a starting point, we developed a three-level classification for ML device autonomy based on how clinical tasks are divided between clinician and ML device (lowest to highest; figure 3).

1. Assisted devices are characterised by overlap in what clinician and device contribute to the task, but where clinicians provide the decision on the task. Such overlap or duplication occurs when clinicians need to confirm or approve device provided information or decisions.

2. Autonomous information is characterised by a separation between what device and clinician contribute to the task, where devices contribute information that clinicians can use to make decisions.

3. Autonomous decision is where device provides the decision for the clinical task which can then be enacted by clinicians or the device itself.

Conceptually, there is also a zero level, representing the complete absence of automation where clinicians perform tasks manually without any device assistance.

Two investigators independently assessed the stage of automation and level of autonomy (DL and FM). Interrater agreement was assessed using absolute agreement, two-way mixed effects intraclass correlation coefficient (ICC). Agreement for stage of automation was ICC=0.7

(95% CI 0.53 to 0.82) indicating moderate to good agreement and for level of autonomy was ICC=0.97 (95% CI 0.95 to 0.98) indicating excellent agreement.²² Disagreements were resolved by consensus. A narrative synthesis then integrated findings into descriptive summaries for each category of ML devices.

RESULTS

Fifty-nine FDA approvals for ML devices met the inclusion criteria covering 49 unique devices (table 1). Six devices had two approvals and two had three approvals.

FDA approvals

The earliest approval was in 2008 for IB Neuro²³ which produces perfusion maps and quantification of blood volume and flow from brain MRI. However, the majority of approvals were observed in recent years (2016=3; 2017=5; 2018=22; 2019=27; 2020=2).

Most approvals (n=51) were via premarket notification (PMN) for devices which are substantially equivalent to existing and legally marketed devices. Only two were via premarket approval (PMA), the most stringent pathway involving regulatory and scientific review, including clinical trials to evaluate safety and efficacy.¹³ The remaining six approvals were De Novo classification; a less onerous alternative to PMA for low to moderate risk devices where there is no substantially equivalent predicate. All PMN and De Novo approvals (n=57) were for class 2 devices, while both pPMAs (n=2) were for class 3 devices, which are classified as moderate and high levels of risk, respectively.

Clinical tasks and diseases supported by ML devices

We identified five distinct clinical tasks supported by ML devices. Most (n=35) assisted with diagnostic tasks assisting with the detection, identification or assessment of disease, or risk factors, such as breast density. The second most common were triage tasks (n=10), where devices assisted with prioritising cases for clinician review, by flagging or notifying cases with suspected positive findings of time-sensitive conditions, such as stroke. Less common tasks were medical procedures (n=2), where devices either assisted users performing diagnostic or interventional procedures. Treatment tasks (n=1) where devices provided CDS recommendations for changes to therapy regimes. Monitoring tasks (n=1) involved devices assisting clinicians to monitor patient trajectory over time.

Twenty-three devices were indicated for a specific disease, and nine could be reasonably associated with a disease. The most common diseases were cancers, especially of the breast, lung, liver and prostate (table 2). Others were stroke (intracranial haemorrhage and large vessel occlusion) and heart diseases. Two devices were indicated for two separate diseases.^{24 25} The remaining 17 devices were indicated for applications broader than a specific disease.

Device inputs and outputs

The majority of devices used image data (n=42), these included computed tomography (CT; n=15), magnetic

resonance imaging (MRI; n=10), X-ray (n=5), digital breast tomosynthesis (n=3), digital mammography (n=3), echocardiography (n=3), fluoroscopy (n=1), fundus imaging (n=1), optical coherence tomography (n=1), positron emission tomography (PET; n=1) and ultrasound (n=1).

The remaining seven used signal data. These included, electrocardiography (n=3), phonocardiography (n=2), polysomnography (n=1), blood glucose and insulin pump data (n=1) and biometric data from wearables (n=1).

We identified nine common means by which ML devices communicated results (table 3).

ML method

Manufacturers descriptions of ML method were varied. Most described a family of techniques (ML=14; deep learning=11), followed by generic descriptors (AI=15). Specific ML techniques were the least frequently reported (convolutional neural network=6; neural network=1; deep neural network=1; deep convolution neural network=1).

Stage of decision-making automated or assisted by ml devices

Most devices aided information analysis (n=14) and decision selection (n=29). ML devices also, but less commonly, aided in information acquisition (n=4) and action implementation (n=2), the earliest and latest stages of decision making, respectively.

Information acquisition

None of the devices acquired information, but instead aided presentation by enhancing the quality of CT, MRI and PET images²⁶⁻²⁹ thereby assisting clinician interpretation. One representative device, SubtleMR²⁸ reduces noise and increases image sharpness of head, neck, spine and knee MRI scans. SubtleMR receives DICOM (Digital Imaging and Communications in Medicine) image data from and returns enhanced DICOM images to a PACS (picture archiving and communication system) server.

Information analysis

Information analysis provides clinicians with new information derived from processing raw inputs. Devices provided analysis in the form of quantification³⁰⁻³⁷ or automatic coding of features or events.^{38 39} For example, IcoBrain³¹ provides volumetric quantification of brain structures from MRI or CT scans, which can aid in the assessment of dementia and traumatic brain injury, while EnsoSleep³⁹ automatically codes events in sleep studies such as stages of sleep and obstructive apnoeas to assist with the diagnosis of sleep disorders.

Decision selection

Decision selection provides a decision that is an outcome for the clinical task, such as triage notifications,^{25 40-48} case level findings of disease,^{24 49-53} identification of features indicative of disease,⁵⁴⁻⁵⁹ or clinical classifications or gradings.⁶⁰⁻⁶⁴ One device providing triage notifications is Briefcase.^{25 40 41} Briefcase assists radiologists triage

Table 1 Characteristics of ML medical devices approved by the US Food and Drug Administration (2008–2020)

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
Information acquisition													
K182901 ²⁶	Aquilion Precision (TSX-304A/1 and /2) V8.8 with AICE	Canon Medical Systems	PMN	July 2019	2	Diagnosis	Improve image quality and reduce noise for displaying cross-sectional volumes of the whole body.	CT	Enhanced images	DCNN	No	Information	
K193170 ²⁷	Deep Learning Image Reconstruction	GE Healthcare Japan	PMN	December 2019	2	Diagnosis	Computer reconstruction of images to produce cross-sectional images of the head and whole body.	CT	Enhanced images	DNN	No	Information	
K191688 ²⁸	SubtleMR	Subtle Medical	PMN	September 2019	2	Diagnosis	Reduce image noise for head, spine, neck, and knee MRI or increase image sharpness for head MRI.	MRI	Enhanced images	GNN	Yes	Information	
K182336 ²⁹	SubtlePET	Subtle Medical	PMN	November 2018	2	Diagnosis	Reduce noise to increase image quality.	PET	Enhanced images	GNN	Yes	Information	
Information analysis													
K182616 ³⁸	NvisionVLE Imaging System	NinePoint Medical	PMN	November 2018	2	Diagnosis	Enhanced visualisation with colourised display of common image features of human tissue.	Optical coherence tomography	Coding features or events	DL	No	Information	
K162627 ³⁹	EnsoSleep	EnsoData	PMN	March 2017	2	Diagnosis	Automatically score sleep studies.	Polysomnography	Coding features or events	AI	Yes	Assistive	Sleep and respiratory related sleep disorders (i)
K173780 ³⁰	EchoMD Automated Ejection Fraction Software	Bay Labs	PMN	June 2018	2	Diagnosis	Automated estimation of left ventricular ejection fraction.	Echocardiogram	Quantification	GNN	Yes	Information	
K192130 ³¹	IcoBrain	Icometrix NV	PMN	December 2019	2	Diagnosis	Automatic labelling, visualisation, and volumetric quantification of brain structures.	CT / MRI	Quantification	AI	Yes	Information	

Continued



Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K080762 ²³	IB Neuro	Imaging Biometrics	PMN	May 2008	2	Diagnosis	Generate parametric perfusion maps of Relative Cerebral Blood Volume, Cerebral Blood Flow, Mean Transit Time, and Time to Peak.	MRI	Quantification	ML	Yes	Information	
K183268 ³²	AI-Rad Companion (Cardiovascular)	Siemens Medical Solutions USA	PMN	September 2019	2	Diagnosis	Quantification of heart volume, calcium volume in coronary arteries, and maximum aorta diameters at typical landmarks.	CT	Quantification	DL	Yes	Information	Cardiovascular diseases (i)
K183271 ⁶⁹	AI-Rad Companion (Pulmonary)	Siemens Medical Solutions USA	PMN	July 2019	2	Diagnosis	Quantification of lungs and lung lobes, and specified lung lesions.	CT	Quantification	DL	Yes	Information	disease of the lungs, Lung cancer (i)
K161322 ⁷⁰	CT CoPlot	ZepMed	PMN	December 2016	2	Diagnosis	Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures.	CT	Quantification	ML	Yes	Information	
K163253 ³³	Arterys Cardio DL Arterys	Arterys	PMN	January 2017	2	Diagnosis	Quantification of the blood flow to the heart and its major vessels.	MRI	Quantification	DL	Yes	Assistive	
K183282 ⁶⁸	Biovitals Analytics Engine	Biofourmis Singapore	PMN	August 2019	2	Monitoring	Calculates time series Biovitals Index which reflects changes in the patient's vital signs from their measured baseline.	Continuous biometric data from worn sensors	Quantification	ML	Yes	Assistive	
K191278 ³⁴	RSI-MRI+	HealthLytix	PMN	November 2019	2	Diagnosis	Automatic prostate segmentation, quantification, and reporting.	MRI	Quantification	DL	Yes	Assistive	Prostate cancer (a)

Continued

Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K173939 ³⁵	Quantib Brain	Quantib B.V.	PMN	March 2018	2	Diagnosis	Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures for GE Advantage Workstation or AW Server.	MRI	Quantification	ML	No	Assistive	
K182564 ³⁶	Quantib ND	Quantib B.V.	PMN	December 2018	2	Diagnosis	Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures for Myrian.	MRI	Quantification	ML	Yes	Assistive	
K191171 ³⁷	EchoGo Core	Ultramics	PMN	November 2019	2	Diagnosis	Automatically measures standard cardiac parameters including Ejection Fraction, Global Longitudinal Strain, and Left Ventricular volume.	Echocardiogram	Quantification	ML	Yes	Assistive	
Decision selection													
DEN180001 ⁴⁸	Idx-DR	IDX	De novo	April 2018	2	Diagnosis	Automatically detect more than mild diabetic retinopathy in adults with diabetes.	Fundus images	Case level finding of disease	CNN	Yes	Decision	Diabetic retinopathy (i)
K181988 ⁵⁰	eMurmur ID	CSD Labs	PMN	April 2019	2	Diagnosis	Automated analysis of specific heart sounds, including S1, S2, physiological and pathological heart murmurs.	Phonocardiogram	Case level finding of disease	ML	Yes	Assistive	Heart murmur (i)
K192004 ²⁴	Eko Analysis Software	Eko Devices	PMN	January 2020	2	Diagnosis	Detects suspected heart murmurs, atrial fibrillation, and normal sinus rhythm. Calculates heart rate, QRS duration and EMAT.	Phonocardiogram & ECG	Case level finding of disease	NIN	Yes	Assistive	Cardiac Arrhythmias, Heart murmur (i)

Continued



Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K192109 ⁵¹	KOALA	IB Lab	PMN	November 2019	2	Diagnosis	Metric measurements of the joint space width and indicators for presence of osteoarthritis on PA/AP knee X-ray images.	X-ray	Case level finding of disease	ML	Yes	Assistive	Knee Osteoarthritis (i)
DEN180005 ⁵²	OsteoDetect	Imagen Technologies	De novo	May 2018	2	Diagnosis	Identify and highlight distal radius fractures during the review of posterior-anterior and lateral radiographs of adult wrists.	X-ray	Case level finding of disease	DL	Yes	Assistive	Wrist fracture (i)
K180432 ⁵³	AI-ECG Platform	Shenzhen Carewell Electronics	PMN	November 2018	2	Diagnosis	Interprets ECG for cardiac abnormalities, such as arrhythmias, myocardial infarction, ventricular hypertrophy, and abnormal ST-T changes.	ECG	Case level finding of disease	AI	Yes	Assistive	Cardiac Arrhythmias (i)
K180125 ⁶⁰	Powerlook Density Assessment Software	iCad	PMN	April 2018	2	Diagnosis	Analyses the dense tissue area of each breast, providing a Category of 1–4 consistent with ACR BI-RADS fifth edition.	Digital breast tomosynthesis	Clinical grading or scoring	ML	Yes	Decision	Breast cancer (a)
K172983 ⁶¹	HealthCCS	Zebra Medical Vision	PMN	June 2018	2	Diagnosis	Quantifies calcification in the coronary arteries reporting a risk category.	CT	Clinical grading or scoring	GNN	Yes	Information	Coronary artery disease (i)
K173542 ⁶²	Arterys Oncology DL	Arterys	PMN	January 2018	2	Diagnosis	Quantification of lung nodules and liver lesions, reporting in accordance with Lung-RADS and LI-RADS.	CT / MRI	Clinical grading or scoring	DL	Yes	Assistive	Lung and liver cancers. (a)

Continued

Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K170540 ⁵³	DM-Density	Densitas	PMN	February 2018	2	Diagnosis	Analyses and reports breast density in accordance with the BI-RADS density classification scales.	Digital mammography	Clinical grading or scoring	AI	Yes	Assistive	Breast cancer (a)
K190442 ⁵⁴	Koios DS for Breast	Koios Medical	PMN	July 2019	2	Diagnosis	Automatically classify user-selected region(s) of interest containing a breast lesion into four BI-RADS-aligned categories.	Ultrasound	Clinical grading or scoring	ML	Yes	Assistive	Breast cancer (i)
K170568 ⁵⁴	Cardiologs	Cardiologs Technologies	PMN	June 2017	2	Diagnosis	Identification and labeling abnormal cardiac rhythms.	ECG	Identify features of disease	DL	Yes	Assistive	Cardiac Arrhythmias (i)
P080003/S008 ⁵⁵	Selenia Dimensions 3D System High Resolution Tomosynthesis	Hologic	PMA	October 2019	3	Diagnosis	Identification of clinically relevant regions of interest for breast screening and diagnosis.	Digital breast tomosynthesis	Identify features of disease	AI	No	Assistive	Breast cancer (i)
K191994 ⁵⁶	Profound AI	iCad	PMN	October 2019	2	Diagnosis	Detect malignant soft tissue densities and calcifications in breast images.	Digital breast tomosynthesis	Identify features of disease	DL	Yes	Assistive	Breast cancer (a)
DEN170022 ⁵⁷	Quantx	Quantitative Insights	De novo	July 2017	2	Diagnosis	Analysis of selected regions, providing comparison with similar lesions with a known ground truth.	MRI	Identify features of disease	AI	Yes	Assistive	Breast cancer (a)
K161201 ⁵⁸	ClearRead CT	Riverain Technologies	PMN	September 2016	2	Diagnosis	Generates secondary vessel suppressed lung images, marking regions of interest to aid in the detection of pulmonary nodules.	CT	Identify features of disease	ML	Yes	Assistive	Lung cancers (a)

Continued

Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K192287 ⁵⁹	Transpara	ScreenPoint Medical BV	PMN	December 2019	2	Diagnosis	Identify regions suspicious for breast cancer and assess their likelihood of malignancy.	Digital mammography	Identify features of disease	ML	Yes	Assistive	Breast cancer (i)
K182218 ⁸⁶	FerriSmart Analysis System	Resonance Health Analysis Services	PMN	November 2018	2	Diagnosis	Measure liver iron concentration.	MRI	Quantification	CNN	Yes	Decision	
K191370 ⁸⁷	DreaMed Advisor Pro	DreaMed Diabetes	PMN	July 2019	2	Treatment	Analyse blood glucose and insulin pump data to generate recommendations for optimising insulin pump dose ratios.	Blood glucose and insulin pump data	Treatment recommendations	ML	Yes	Assistive	Type 1 diabetes (i)
K190896 ⁴⁰	Briefcase (CSF)	AiDoc Medical	PMN	May 2019	2	Triage	Analyse images, notifying cases with suspected positive findings of linear lucencies in the cervical spine bone in patterns compatible with fractures.	CT	Triage notifications	AI	Yes	Decision	Cervical spine fracture (i)
K190072 ²⁵	Briefcase (ICH and PE)	AiDoc Medical	PMN	April 2019	2	Triage	Analyse images, notifying cases with suspected positive findings of Intracranial Haemorrhage or Pulmonary Embolism pathologies.	CT	Triage notifications	AI	Yes	Decision	Intracranial Haemorrhage, Pulmonary Embolism (i)
K192383 ⁴¹	Briefcase (LVO)	AiDoc Medical	PMN	December 2019	2	Triage	Analyse images, notifying cases with suspected positive findings of Large Vessel Occlusion pathologies.	CT	Triage notifications	AI	Yes	Decision	Large Vessel Occlusion (i)
K183285 ⁴²	CmTriage	CureMetrix	PMN	March 2019	2	Triage	Analyse screening mammograms, notifying cases with at least one suspicious finding at the exam level.	Digital mammography	Triage notifications	ML	Yes	Decision	Breast cancer (a)

Continued

Table 1 Continued

Approval no	Device	Manufacturer	Scheme	Date	Class	Clinical task	Device task	Device input	Device output	ML method	SaMD	Autonomy	Disease(s)
K183182 ⁴³	Critical Care Suite	GE Medical Systems	PMN	August 2019	2	Triage	Analyse and notify cases with suspected findings of pneumothorax.	X-ray	Triage notifications	DL	Yes	Decision	Pneumothorax (i)
K182177 ⁴⁴	AccipioX	MaxQ-AI	PMN	October 2018	2	Triage	Identify and notify cases with suspected findings of acute intracranial haemorrhage.	CT	Triage notifications	AI	Yes	Decision	Intracranial Haemorrhage (i)
DEN170073 ⁴⁵	ContaCT	Viz.AI	De novo	February 2018	2	Triage	Identify and notify cases with suspected findings of large vessel occlusion.	CT	Triage notifications	AI	Yes	Decision	Large Vessel Occlusion (i)
K192320 ⁴⁶	HealthCXR	Zebra Medical Vision	PMN	November 2019	2	Triage	Automatically identify and notify suspected findings of pleural effusion.	X-ray	Triage notifications	AI	Yes	Decision	Pleural Effusion (i)
K190424 ⁴⁷	HealthICH	Zebra Medical Vision	PMN	June 2019	2	Triage	Automatically identify and notify suspected findings of Intracranial Haemorrhage.	CT	Triage notifications	AI	Yes	Decision	Intracranial Haemorrhage (i)
K190362 ⁴⁸	HealthPNX	Zebra Medical Vision	PMN	May 2019	2	Triage	Automatically identify and notify suspected findings of Pneumothorax.	X-ray	Triage notifications	AI	Yes	Decision	Pneumothorax (i)
Action Implementation													
K191713 ⁶⁷	FluoroShield	Omega Medical Imaging	PMN	October 2019	2	Procedure	Automated region of interest in fluoroscopic applications.	Fluoroscopy	Automatic control of device	AI	No	Decision	
DEN190040 ⁶⁶	Caption Guidance	Bay Labs	De novo	February 2020	2	Procedure	Real-time guidance during echocardiography acquisition to assist obtaining correct images of standard diagnostic views and orientations.	Echocardiogram	Automatic control of device	DL	Yes	Assistive	

Disease(s): (i) indicated disease; (a) associated disease.

ML method as described by manufacturer.

Autonomy: assistive; Information, autonomous information; Decision, autonomous decision.

AI, artificial intelligence; BI-RADS, Breast Imaging-Reporting and Data System; CNN, convolutional neural network; DCNN, Deep Convolutional Neural Network; DL, deep learning; DNN, deep neural network; LI-RADS, Liver Imaging-Reporting and Data System; ML, machine learning; NN, neural network; PMA, premarket approval; PMN, premarket notification; SaMD, Software as a Medical Device.

Table 2 Diseases indicated or associated with ML devices

Indicated or associated disease*	N	Diagnose	Triage	Treatment
Breast cancer	8	7	1	
Cardiac arrhythmias	3	3		
Intracranial haemorrhage	3		3	
Lung cancer	3	3		
Heart murmurs	2	2		
Large vessel occlusion	2		2	
Pneumothorax	2		2	
Cervical spine fracture	1		1	
Coronary artery disease	1	1		
Diabetic retinopathy	1	1		
Osteoarthritis	1	1		
Liver cancer	1	1		
Pleural effusion	1		1	
Prostate cancer	1	1		
Pulmonary embolism	1		1	
Sleep and respiratory related sleep disorders	1	1		
Diabetes mellitus type 1	1			1
Wrist fracture	1	1		

*Two devices were indicated for two diseases.
ML, machine learning.

time-sensitive cases by flagging and displaying notifications for cases with suspect positive findings of cervical spine fracture,⁴⁰ large vessel occlusion,⁴¹ intracranial haemorrhage and pulmonary embolism²⁵ as they are received. A device providing case level findings of disease

is AI-ECG Platform.⁵³ It reports whether common cardiac conditions are present, such as arrhythmias and myocardial infarction. While clinicians can view the original tracings, the device reports on the entire case. In contrast, a device providing feature level detection of disease is

Table 3 ML device output by type

Output type	Devices	Description
Quantification	13	Quantification of information derived from the images, such as, cardiac function and blood flow, ^{23 30 32 33 37} or volume of structures including the brain, ^{31 35 36 70} and prostate. ³⁴
Triage notifications	10	Triage notification alert clinicians to cases with suspected positive findings. ^{25 40–48}
Case-level finding of disease	6	Case level finding of disease such as, wrist fractures, ⁵² diabetic retinopathy, ⁴⁹ osteoarthritis, ⁵¹ heart murmurs ^{24 50} and cardiac arrhythmias. ^{24 53}
Identify features of disease	6	Identify features of disease thereby drawing clinician attention to them, such as prompting breast ^{55–57 59} or lung ⁵⁸ cancers on images or cardiac arrhythmias on ECG tracings. ⁵⁴
Clinical grading or scoring	5	Clinical grading or scoring (n=5) on standardised clinical assessment instruments, such as BI-RADS, ^{60 63 64} LI-RADS, ⁶² lung-RADS, ⁶² or Agatston-equivalent scores. ⁶¹
Enhanced images	4	Enhanced images with reduced noise and improved image quality. ^{26–29}
Automatic coding of features or events	2	Automatic coding of features or events in the data, such as sleep stages and respiratory events in polysomnography data, ³⁹ or colour coding structures in optical coherence tomography. ³⁸
Automatic control of electronic or mechanical devices	2	Automatic control of electronic or mechanical devices, such as fluoroscope collimator ⁶⁷ and automatic recording of ultrasound clips dependent on detected image quality. ⁶⁶
Treatment recommendations	1	Treatment recommendations, such as adjustments to insulin pump dose ratios. ⁸⁷

BI-RADS, Breast Imaging-Reporting and Data System; LI-RADS, Liver Imaging-Reporting and Data System; ML, machine learning.

Profound AI,⁵⁶ which detects and marks features indicative of breast cancer on digital breast tomosynthesis exams. It is intended to be used concurrently by radiologists while interpreting exams, drawing attention to features which radiologists may confirm or dismiss. A device reporting clinical classifications or grades is DM-Density,⁶³ which reports breast density grading for digital monography cases according to the American Collage of Radiology's Breast Imaging-Reporting and Data System Atlas.⁶⁵

Action implementation

Devices providing action implementation included Caption Guidance⁶⁶ and FluroShield⁶⁷; these implemented decisions through the automatic control of an electronic or mechanical device. Caption Guidance⁶⁶ assists with acquisition of echocardiograms, providing real-time guidance to sonographers and feedback on detected image quality. Ultrasounds are automatically captured when the correct image quality is detected. FluroShield⁶⁷ automatically controls the collimator during the fluoroscopy to provide a live view of a region of interest, with a lower refresh rate of once or twice per second for the wider field of view, thereby reducing radiation exposure to patient and clinician.

ML device autonomy

Nearly half (n=23) of devices were assistive, characterised by indications emphasising clinician responsibility for the final decision or statements limiting the extent to which the device could be relied on (box 1). Assistive devices comprised all devices providing feature level detection,^{54–59} five of six devices reporting a case level finding of disease,^{24 50–53} and almost half of devices providing quantification.^{33–37 68} Notwithstanding clinician responsibility to patients, the indications for the remaining devices did not

specify such limitations, when used as indicated. Consequently, those devices appeared to automate functions otherwise performed by clinicians, to a greater extent than assistive devices. Fourteen devices provide autonomous decisions that clinicians could act on; these were primarily devices providing triage notifications,^{25 40–48} but also included IDx-DR⁴⁹ a device providing case-level findings of diabetic retinopathy, allowing screening in primary practice where results are used as the basis for specialist referral for diagnosis and management. Twelve devices provide autonomous information, that clinicians could use in their decision making to determine an outcome for clinical tasks. These included devices providing enhanced images,^{26–29} quantification^{23 30–32 69 70} and one device which coded features or events.³⁸

DISCUSSION

Main findings and implications

The way that algorithms are embedded in medical devices shapes how clinicians interact with them, with different profiles of risk and benefit. We demonstrate how the stages of automation framework,¹⁹ can be applied to determine the stage of clinician decision making assisted by ML devices. Together with our level of autonomy framework, these methods can be applied to examine how ML algorithms are used in clinical practice, which may assist addressing the dearth of human factors evaluations related to the use of ML devices in clinical practice.¹⁷ Such analyses (table 1) permit insight into how ML devices may impact or change clinical workflows and practices, and how these may impact healthcare delivery.

While FDA approval of ML devices is a recent development, only six approvals in this study were via De Novo classification for new types of medical devices. Most approvals were via the PMN pathway for devices that are substantially equivalent to existing predicate devices. Some predicates could be traced to the ML device De Novo's, while others were non-ML devices with similar indications except using different algorithms. As the FDA assesses all medical devices on the same basis, regardless of ML utilisation, it is unsurprising that ML medical devices largely follow in the footsteps of their non-ML forebears. Most were assistive or autonomous information which left responsibility for clinical decisions to clinicians.

We identified an interesting group of devices, primarily triage devices, which provided autonomous decisions, independent of clinicians. These triage devices appeared to perform tasks intended to supplement clinician workflow, rather than to automate or replace existing clinician tasks. The expected benefit is prioritising the reading of cases with suspected positive findings for time-sensitive conditions, such as stroke, thereby reducing time to intervention, which may improve prognosis. Unlike PMNs, De Novo classifications report more details, including identified risks. The De Novo for the triage device, ContaCT,⁴⁵ identifies risks associated with false-negatives that could

Box 1 Examples of FDA-approved indications specifying responsibility for the final decision on the device task resides with the clinician. For further examples, see online supplemental appendix A

'All automatically scored events are subject to verification by a qualified clinician.'³⁹

'Not intended for making clinical decisions regarding patient treatment or for diagnostic purposes.'⁶⁸

'Intended as an additional input to standard diagnostic pathways and is only to be used by qualified clinicians.'³⁷

'Interpretations offered by (device) are only significant when considered in conjunction with healthcare provider over-read and including all other relevant patient data.'⁵⁰

'Should not be used in lieu of full patient evaluation or solely relied on to make or confirm a diagnosis.'⁵¹

'The clinician retains the ultimate responsibility for making the pertinent diagnosis based on their standard practices.'⁶²

'Patient management decisions should not be made solely on the results.'⁶⁴

'Provides adjunctive information and is not intended to be used without the original CT series.'⁶⁸

lead to incorrect or delayed patient management, while false-positives may deprioritise other cases.

Likewise, the diabetic retinopathy screening device, IDx-DR⁴⁹ appears to supplement existing workflows by permitting screening in primary practice that would otherwise be impossible. The goal is to increase screening rates for diabetic retinopathy by improving access to screening and reducing costs.⁷¹ The De Novo describes risks that false-negatives may delay detection of retinopathy requiring treatment, while false-positives may subject patients to additional and unnecessary follow-up.⁴⁹ However, the device may enable far greater accessibility to regular screening.

In contrast, with assistive devices there is overlap between what the clinician and device does. Despite many of these ML devices providing decision selection, such as reporting on the presence of disease, the approved indications of all assistive devices—nearly half of reviewed devices—emphasised that decisions are the responsibility of the clinician (box 1). Such stipulations specify how device information should be used and may stem from several sources, such as legal requirements for tasks: who can decide what, for example, diagnose or prescribe medicines, and legal liability about who is accountable when things go wrong. However, the trustworthiness of devices cannot be inferred by the presence of such indications.

Assistive devices change how clinicians work and can introduce new risks.⁷² Instead of actively detecting and diagnosing disease, through patient examination, diagnostic imaging or other procedures, the clinician's role is changed by the addition of the ML device as a new source of information. Crucially, indications requiring clinicians to confirm or approve ML device findings create new tasks for clinicians; to provide quality assurance for device results, possibly by scrutinising the same inputs as the ML device, together with consideration of additional information.

The benefit of assistive ML devices is the possibility of detecting something that might have otherwise been missed. However, there is risk that devices might bias clinicians; that is, ML device errors may be accepted as correct by clinicians, resulting in errors that might not have otherwise occurred.^{9 73} Troublingly, people who suffer these automation biases exhibit reduced information seeking^{74–76} and reduced allocation of cognitive resources to process that information,⁷⁷ which in turn reduces their ability to recognise when the decision support they have received is incorrect. While improving ML device accuracy reduces opportunities for automation bias errors, high accuracy is known to increase the rate of automation bias,⁷⁸ likely rendering clinicians less able to detect failures when they occur. Of further concern, is evidence showing far greater performance consequences when later stage automation fails, which is most evident when moving from information analysis to decision selection.⁷⁹ Greater consequences could be due to reduced situational awareness as automation takes over more stages of human information processing.⁷⁹

Indeed, the De Novo for Quantx,⁵⁷ an assistive device which identifies features of breast cancer from MRI, describes the risk of false-negatives which may lead to misdiagnosis and delay intervention, while false-positives may lead to unnecessary procedures. The De Novo for OsteoDetect⁵² likewise identifies a risk of false-negatives that ‘users may rely too heavily on the absence of (device) findings without sufficiently assessing the native image. This may result in missing fractures that may have otherwise been found.’⁵² While false-positives may result in unnecessary follow-up procedures. These describe the two types of automation bias errors which can occur when clinicians act on incorrect CDS. Omission errors where clinicians agree with CDS false-negatives and consequently fail to diagnose a disease, and commission errors whereby clinicians act on CDS false-positives by ordering unnecessary follow-up procedures.^{9 80}

Other risks identified in De Novo classifications^{45 52 57} include device failure, and use of devices on unintended patient populations, with incompatible hardware and for non-indicated uses. Such risks could result in devices providing inaccurate or no CDS. Controls outlined in De Novos focused on software verification and validation, and labelling, to mitigate risks of device and user errors, respectively.

These findings have several implications. For clinicians, use of ML devices needs to be consistent with labelling and results scrutinised according to clinicians' expertise and experience. Scrutiny of results is especially critical with assistive devices. There needs to be awareness of the potential for ML device provided information to bias decision-making. Clinicians also need to be supported to work effectively with ML devices, with the training and resources necessary to make informed decisions about use and how to evaluate device results. For ML device manufacturers and implementers, the choice of how to support clinicians is important, especially the choice of which tasks to support, what information to provide and how clinicians will integrate and use those devices within their work. For regulators, understanding the stage and extent of human information processing automated by ML devices may complement existing risk categorisation frameworks,^{81 82} by accounting for how the ML device contribution to decision-making modifies risk for the intended use of device provided information; to treat or diagnose, to drive clinical management or to inform clinical management.⁸¹ Regulators could improve their reporting of ML methods used to develop the algorithms utilised by devices. These algorithms are akin to the ‘active ingredient’ in medicines as they are responsible for the device's action. However, consistent with the previous study we found that the public reporting of ML methods varied considerably but was generally opaque and lacking in detail.¹² Presently, the FDA only approves devices with ‘locked’ algorithms,⁸² but are moving towards a framework that would permit ML devices which learn and adapt to real-world data.⁸³ Such a framework is expected to involve precertification of vendors and submission

of algorithm change protocols.⁸² It will be important to continually evaluate the clinician-ML device interactions which may change with regulatory frameworks.

Finally, there are important questions about responsibility for ML device provided information and the extent to which clinicians should be able to rely on it. While exploration of these questions exceeds the scope of this article, models of use that require clinicians to double check ML devices results, may be less helpful than devices whose output can be acted on. As ML devices become more common there needs to be clearly articulated guidelines on the division of labour between clinician and ML devices, especially in terms of who is responsible for which decisions and under what circumstances. In addition to the configuration of tasks between clinician and ML devices, how devices work and communicate with clinicians is crucial and requires further study. The ability of ML devices to explain decisions through presentation of information, such as marking suspected cancers on images or using explainable AI techniques⁸⁴ will impact how clinicians will assess and make decisions based on ML device provided information.

Limitations

There are several limitations. First, it was not possible to directly search FDA approval databases, the primary source of approvals. Second, the reporting in approvals varied considerably with nearly one third of included approvals not describing ML utilisation. Indeed, all disagreements on device selection occurred where evidence had to be sought from the manufacturer's website and non-peer reviewed sources, where one reviewer located key information the other did not. Consequently, it is possible that some devices may have been missed. Nevertheless, the review provides useful insights in the absence of capability to systematically search primary sources. Our analysis focused on intended use as described in approvals, rather than actual use in the real world, which may differ. Finally, the focus on medical devices limits the review to ML algorithms approved by the FDA. Nevertheless, our methods to examine the stage of human information processing automated and level of autonomy can be applied to examine clinician interaction with the vast majority of ML CDS which are not regulated as medical devices. Indeed, there is an urgent need to ensure ML based CDS are implemented safely and effectively in clinical settings.⁸⁵

CONCLUSION

Our analysis demonstrates the variety of ways in which ML algorithms are embedded in medical devices to support clinicians, the task supported and information provided. Leveraging the benefits of ML algorithms for CDS and mitigating risks, requires a solid working relationship between clinician and the CDS. Such a relationship must be carefully designed, considering how algorithms are embedded in devices, the clinical tasks they support, the

information they provide and how clinicians will interact with them.

Twitter David Lyell @David_Lyell, Enrico Coiera @EnricoCoiera and Farah Magrabi @farahmagrabi

Acknowledgements We wish to acknowledge the invaluable contributions of Didi Surian, Ying Wang and Rhonda Siu.

Contributors DL conceived this study, and designed and conducted the analysis with advice and input from FM and EC. PS and DL screened ML devices for inclusion and performed data extraction. JC identified additional ML devices and screened devices for inclusion. FM and DL assessed the stage of automation and level of autonomy. DL drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

Funding NHMRC Centre for Research Excellence (CRE) in Digital Health (APP1134919) and a Macquarie University Safety Net grant.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information. All data relevant to the analysis are reported in the article. The FDA approval documents analysed are cited in the reference list.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

David Lyell <http://orcid.org/0000-0002-2695-0368>

REFERENCES

- Coiera E. *Guide to health informatics*. Third edition. ed. Boca Raton: CRC Press, Taylor & Francis Group, 2015.
- Matheny M, Israni ST, Ahmed M. Artificial intelligence in health care: the hope, the hype, the promise, the peril. *Natl Acad Med* 2020;94-7.
- Domingos P. *The master algorithm: how the quest for the ultimate learning machine will remake our world*. 1st ed. New York: Basic Books, a member of the Perseus Books Group, 2018.
- Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Hum Factors* 1997;39:230-53.
- Coiera E. The last mile: where artificial intelligence meets reality. *J Med Internet Res* 2019;21:e16323.
- Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231-7.
- Rajkomar A, Hardt M, Howell MD, et al. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;169:866-72.
- Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc* 2017;24:246-50.
- Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc* 2017;24:423-31.
- Federal Food, Drug, and Cosmetic Act, Section 201(h). United States of America.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.
- Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020;3:118.

- 13 Maisel WH. Medical device regulation: an introduction for the practicing physician. *Ann Intern Med* 2004;140:296–302.
- 14 Coiera E. On algorithms, machines, and medicine. *Lancet Oncol* 2019;20:166–7.
- 15 Liu X, Faes L, Kale AU, *et al.* A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1:e271–97.
- 16 Houssami N, Kirkpatrick-Jones G, Noguchi N, *et al.* Artificial intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. *Expert Rev Med Devices* 2019;16:351–62.
- 17 Sujan M, Furniss D, Grundy K, *et al.* Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inform* 2019;26:e100081.
- 18 Ben-Israel D, Jacobs WB, Casha S, *et al.* The impact of machine learning on patient care: a systematic review. *Artif Intell Med* 2020;103:101785.
- 19 Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society* 2000;30:286–97.
- 20 SAE International. *Taxonomy and definitions for terms related to driving automation systems for On-Road motor vehicles: SAE international*, 2016.
- 21 Wickens CD, Mavor A, Parasuraman R. *The future of air traffic control: human operators and automation*. Washington, DC: National Academies Press, 1998.
- 22 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
- 23 U.S. Food and Drug Administration. K080762, IB NEURO, VERSION 1.0, IMAGING BIOMETRICS, LLC. *510(k) Premarket Notification*, 2008. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K080762>
- 24 U.S. Food and Drug Administration. K192004, Eko Analysis Software, Eko Devices Inc. *510(k) Premarket Notification*, 2020. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192004>
- 25 U.S. Food and Drug Administration. K190072, Briefcase (ICH and PE), Aidoc Medical, Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K190072>
- 26 U.S. Food and Drug Administration. K182901, Aquilion Precision (TSX-304A/1 and /2) V8.8 with AiCE, Canon Medical Systems Corporation. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182901>
- 27 U.S. Food and Drug Administration. K193170, Deep Learning Image Reconstruction, GE Healthcare Japan Corporation. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K193170>
- 28 U.S. Food and Drug Administration. K191688, SubtleMR, Subtle Medical, Inc. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K191688>
- 29 U.S. Food and Drug Administration. K182336, SubtlePET, Subtle Medical, Inc. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182336>
- 30 U.S. Food and Drug Administration. K173780, EchoMD Automated Ejection Fraction Software, Bay Labs, Inc. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K173780>
- 31 U.S. Food and Drug Administration. K192130, Icobrain, Icometrix NV. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192130>
- 32 U.S. Food and Drug Administration. K183268, AI-Rad Companion (Cardiovascular), Siemens Medical Solutions USA, Inc. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183268>
- 33 U.S. Food and Drug Administration. K163253, Arterys Cardio DL, ARTERYS INC. *510(k) Premarket Notification*, 2017. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K163253>
- 34 U.S. Food and Drug Administration. K191278, RSI-MRI+, MultiModal Imaging Services Corporation. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K191278>
- 35 U.S. Food and Drug Administration. K173939, Quantib Brain, Quantib BV. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K173939>
- 36 U.S. Food and Drug Administration. K182564, Quantib ND, Quantib BV. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182564>
- 37 U.S. Food and Drug Administration. K191171, EchoGo Core, Ultromics Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K191171>
- 38 U.S. Food and Drug Administration. K182616, NvisionVLE Imaging System, NvisionVLE Optical Probe, NvisionVLE Inflation System, NinePoint Medical, Inc. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182616>
- 39 U.S. Food and Drug Administration. K162627, EnsoSleep, EnsoData, Inc. *510(k) Premarket Notification*, 2017. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K162627>
- 40 U.S. Food and Drug Administration. K190896, BriefCase (CSF), Aidoc Medical, Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K190896>
- 41 U.S. Food and Drug Administration. K192383, Briefcase (LVO), Aidoc Medical, Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192383>
- 42 U.S. Food and Drug Administration. K183285, CmTriage, CureMetrix, Inc. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183285>
- 43 U.S. Food and Drug Administration. K183182, Critical Care Suite, GE Medical Systems, LLC. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183182>
- 44 U.S. Food and Drug Administration. K182177, Accipiolx, MaxQ-AI Ltd. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182177>
- 45 U.S. Food and Drug Administration. DEN170073, ContaCT, Viz.AI, Inc. *Device Classification Under Section 513(f)(2)(De Novo)*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN170073>
- 46 U.S. Food and Drug Administration. K192320, HealthCXR, Zebra Medical Vision, Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192320>
- 47 U.S. Food and Drug Administration. K190424, HealthICH, Zebra Medical Vision Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K190424>
- 48 U.S. Food and Drug Administration. K190362, HealthPNX, Zebra Medical Vision Ltd. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K190362>
- 49 U.S. Food and Drug Administration. DEN180001, IDx-DR, IDx, LLC. *Device Classification Under Section 513(f)(2)(De Novo)*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN180001>
- 50 U.S. Food and Drug Administration. K181988, eMurmur ID, CSD Labs GmbH. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K181988>
- 51 U.S. Food and Drug Administration. K192109, KOALA, IB Lab GmbH. *510(k) Premarket Notification*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K192109>
- 52 U.S. Food and Drug Administration. DEN180005, OsteoDetect, Imagen Technologies, Inc. *Device Classification Under Section 513(f)(2)(De Novo)*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN180005>
- 53 U.S. Food and Drug Administration. K180432, AI-EKG Platform, Shenzhen Carewell Electronics., Ltd. *510(k) Premarket Notification*, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K180432>
- 54 U.S. Food and Drug Administration. K170568, CardioLogs ECG Analysis Platform, Cardiologs Technologies. *510(k) Premarket Notification*, 2017. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K170568>

- 55 U.S. Food and Drug Administration. P080003/S008, Selenia Dimensions 3D System High Resolution Tomosynthesis, HOLOGIC, INC. *Premarket Approval (PMA)*, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P080003S008>
- 56 U.S. Food and Drug Administration. K191994, ProFound AI Software V2.1, iCAD Inc. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K191994>
- 57 U.S. Food and Drug Administration. DEN170022, QuantX, Quantitative Insights, Inc. *Device Classification Under Section 513(f)(2)(De Novo)*, 2017. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?ID=DEN170022>
- 58 U.S. Food and Drug Administration. K161201, ClearRead CT, RIVERAIN TECHNOLOGIES, LLC. 510(k) Premarket Notification, 2016. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K161201>
- 59 U.S. Food and Drug Administration. K192287, Transpara, Screenpoint Medical B.V. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K192287>
- 60 U.S. Food and Drug Administration. K180125, PowerLook Density Assessment Software, iCAD, Inc. 510(k) Premarket Notification, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K180125>
- 61 U.S. Food and Drug Administration. K172983, HealthCCS, Zebra Medical Vision Ltd. 510(k) Premarket Notification, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K172983>
- 62 U.S. Food and Drug Administration. K173542, Arterys Oncology DL, Arterys Inc. 510(k) Premarket Notification, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K173542>
- 63 U.S. Food and Drug Administration. K170540, DM-Density, Densitas, Inc. 510(k) Premarket Notification, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K170540>
- 64 U.S. Food and Drug Administration. K190442, Koios DS for Breast, Koios Medical, Inc. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K190442>
- 65 Sickles E, D'Orsi C, Bassett L. Mammography. In: D'Orsi C, Sickles E, Mendelson E, *et al*, eds. *ACR BI-RADS atlas, breast imaging reporting and data system*. Reston, VA: American College of Radiology, 2013.
- 66 U.S. Food and Drug Administration. DEN190040, Caption Guidance, Bay Labs, Inc. *Device Classification Under Section 513(f)(2)(De Novo)*, 2020. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN190040>
- 67 U.S. Food and Drug Administration. K191713, CS-Series-FP Radiographic / Fluoroscopic Systems With Optional CA-100S / FluoroShield (ROI Assembly), Omega Medical Imaging, LLC. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K191713>
- 68 U.S. Food and Drug Administration. K183282, Biovitals Analytics Engine, Biofourmis Singapore Pte. Ltd. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183282>
- 69 U.S. Food and Drug Administration. K183271, AI-Rad companion (pulmonary), Siemens medical solutions USA, Inc. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183271>
- 70 U.S. Food and Drug Administration. K161322, CT CoPilot, ZEPMED, LLC. 510(k) Premarket Notification, 2016. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K161322>
- 71 Abràmoff MD, Lavin PT, Birch M, *et al*. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
- 72 Coiera E, Westbrook J, Wyatt J. The safety and quality of decision support systems. *Yearb Med Inform* 2006;20–5.
- 73 Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 2012;19:121–7.
- 74 Bagheri N, Jamieson GA. Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In: Vincenzi DA, Mouloua M, Hancock PA, eds. *Human performance, situation awareness and automation: current research and trends*. Mahwah: Lawrence Erlbaum Associates, 2004: 2. 54–9.
- 75 Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision AIDS: the impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making* 2012;6:57–87.
- 76 Lyell D, Magrabi F, Coiera E. Reduced verification of medication alerts increases prescribing errors. *Appl Clin Inform* 2019;10:066–76.
- 77 Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Hum Factors* 2018;60:1008–21.
- 78 Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science* 2007;8:321–48.
- 79 Onnasch L, Wickens CD, Li H, *et al*. Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Hum Factors* 2014;56:476–88.
- 80 Mosier KL, Skitka LJ. Human decision makers and automated decision aids: Made for each other. In: Parasuraman R, Mouloua M, eds. *Automation and human performance: theory and applications*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 1996: 201–20.
- 81 International Medical Device Regulators Forum. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations, 2014.
- 82 U.S. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback*, 2019.
- 83 U.S. Food and Drug Administration. *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*, 2021.
- 84 Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform* 2021;113:103655.
- 85 Petersen C, Smith J, Freimuth RR, *et al*. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an Amia position paper. *J Am Med Inform Assoc* 2021;28:677–84. doi:10.1093/jamia/ocaa319
- 86 U.S. Food and Drug Administration. K182218, FerriSmart Analysis System, Resonance Health Analysis Services Pty Ltd. 510(k) Premarket Notification, 2018. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K182218>
- 87 U.S. Food and Drug Administration. K191370, DreaMed Advisor Pro, DreaMed Diabetes Ltd. 510(k) Premarket Notification, 2019. Available: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K191370>