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

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, such as detecting disease in an image, diagnosis and recommending treatments, have the potential to improve healthcare delivery and patient outcomes. Machine learning (ML) refers more specifically to AI methods that can learn from data. The 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 diagnose or treat human disease automate clinical tasks otherwise done by clinicians. Importantly, CDS operates within a human–technology system, 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.\textsuperscript{2} 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.\textsuperscript{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 devices\textsuperscript{10} and subject to regulation. Increasing numbers of devices that embody ML algorithms are being approved by the US Food and Drug Administration (FDA).\textsuperscript{11,12} Approval requires compliance with standards, as well as evaluation of device safety and efficacy.\textsuperscript{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.\textsuperscript{11,14-16} with little focus on how ML is integrated into clinical practice and the human factors related to their use.\textsuperscript{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.\textsuperscript{18}

While one recent study has described the general characteristics of 64 ML-based medical devices approved by the FDA\textsuperscript{19} 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.\textsuperscript{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).\textsuperscript{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 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).\textsuperscript{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
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. We classified the highest stage of human information processing framework. The device task was examined using the stages of automation based on how clinical tasks are divided between clinician and device. 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. Assistive 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). Inter-rater agreement was assessed using absolute agreement, two-way mixed effects intraclass correlation coefficient (ICC). Agreement for stage of automation was ICC=0.7

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. 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, and computer-based automation, 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. Assistive 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). Inter-rater agreement was assessed using absolute agreement, two-way mixed effects intraclass correlation coefficient (ICC). Agreement for stage of automation was ICC=0.7

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

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

Figure 3 Level of autonomy showing the relationship between clinician and device.
(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. 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 PMNs 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 images 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 quantification or automatic coding of features or events. 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, case level findings of disease, identification of features indicative of disease, or clinical classifications or gradings. One device providing triage notifications is Briefcase. Briefcase assists radiologists triage...
# Table 1 Characteristics of ML medical devices approved by the US Food and Drug Administration (2008–2020)

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

Continued
<table>
<thead>
<tr>
<th>Approval no</th>
<th>Device</th>
<th>Manufacturer</th>
<th>Scheme</th>
<th>Date</th>
<th>Class</th>
<th>Clinical task</th>
<th>Device task</th>
<th>Device input</th>
<th>Device output</th>
<th>ML method</th>
<th>SaMD</th>
<th>Autonomy</th>
<th>Disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K08076223</td>
<td>IB Neuro</td>
<td>Imaging Biometrics</td>
<td>PMN</td>
<td>May 2008</td>
<td>2</td>
<td>Diagnosis</td>
<td>Generate parametric perfusion maps of Relative Cerebral Blood Volume, Cerebral Blood Flow, Mean Transit Time, and Time to Peak.</td>
<td>MRI</td>
<td>Quantification</td>
<td>ML</td>
<td>Yes</td>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>K18326832</td>
<td>Al-Rad Companion (Cardiovascular)</td>
<td>Siemens Medical Solutions USA</td>
<td>PMN</td>
<td>September 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Quantification of heart volume, calcium volume in coronary arteries, and maximum aorta diameters at typical landmarks.</td>
<td>CT</td>
<td>Quantification</td>
<td>DL</td>
<td>Yes</td>
<td>Information</td>
<td>Cardiovascular diseases (i)</td>
</tr>
<tr>
<td>K18327149</td>
<td>Al-Rad Companion (Pulmonary)</td>
<td>Siemens Medical Solutions USA</td>
<td>PMN</td>
<td>July 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Quantification of lungs and lung lobes, and specified lung lesions.</td>
<td>CT</td>
<td>Quantification</td>
<td>DL</td>
<td>Yes</td>
<td>Information</td>
<td>disease of the lungs, Lung cancer (i)</td>
</tr>
<tr>
<td>K16132270</td>
<td>CT CoPilot</td>
<td>ZepMed</td>
<td>PMN</td>
<td>December 2016</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures.</td>
<td>CT</td>
<td>Quantification</td>
<td>ML</td>
<td>Yes</td>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>K16325333</td>
<td>Arterys Cardio DL</td>
<td>Arterys</td>
<td>PMN</td>
<td>January 2017</td>
<td>2</td>
<td>Diagnosis</td>
<td>Quantification of the blood flow to the heart and its major vessels.</td>
<td>MRI</td>
<td>Quantification</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td></td>
</tr>
<tr>
<td>K18328268</td>
<td>Biovitals Analytics Engine</td>
<td>Biofourmis Singapore</td>
<td>PMN</td>
<td>August 2019</td>
<td>2</td>
<td>Monitoring</td>
<td>Calculates time series Biovitals Index which reflects changes in the patient’s vital signs from their measured baseline.</td>
<td>Continuous biometric data from worn sensors</td>
<td>Quantification</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td></td>
</tr>
<tr>
<td>K19127834</td>
<td>RSI-MRI+</td>
<td>HealthLytx</td>
<td>PMN</td>
<td>November 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatic prostate segmentation, quantification, and reporting.</td>
<td>MRI</td>
<td>Quantification</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td>Prostate cancer (a)</td>
</tr>
</tbody>
</table>
### Table 1

**Decision selection**

<table>
<thead>
<tr>
<th>Approval no</th>
<th>Device</th>
<th>Manufacturer</th>
<th>Scheme</th>
<th>Date</th>
<th>Class</th>
<th>Clinical task</th>
<th>Device task</th>
<th>Device input</th>
<th>Device output</th>
<th>ML method</th>
<th>SaMD</th>
<th>Autonomy</th>
<th>Disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1739396⁵⁶</td>
<td>Quantib Brain</td>
<td>Quantib B.V.</td>
<td>PMN</td>
<td>March 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures for GE Advantage Workstation or AW Server.</td>
<td>MRI</td>
<td>Quantification</td>
<td>ML</td>
<td>No</td>
<td>Assistive</td>
<td></td>
</tr>
<tr>
<td>K1825646⁵⁶</td>
<td>Quantib ND</td>
<td>Quantib B.V.</td>
<td>PMN</td>
<td>December 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatic labelling, visualisation, and volumetric quantification of segmentable brain structures for Myrian.</td>
<td>MRI</td>
<td>Quantification</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td></td>
</tr>
<tr>
<td>K191171⁵⁷</td>
<td>EchoGo Core</td>
<td>Ultromics</td>
<td>PMN</td>
<td>November 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatically measures standard cardiac parameters including Ejection Fraction, Global Longitudinal Strain, and Left Ventricular volume.</td>
<td>Echocardiogram</td>
<td>Quantification</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td></td>
</tr>
<tr>
<td>DEN180001⁵⁹</td>
<td>Idx-DR</td>
<td>IDX</td>
<td>De novo</td>
<td>April 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatically detect more than mild diabetic retinopathy in adults with diabetes.</td>
<td>Fundus images</td>
<td>Case level finding of disease</td>
<td>CNN</td>
<td>Yes</td>
<td>Decision</td>
<td>Diabetic retinopathy (i)</td>
</tr>
<tr>
<td>K181988⁵⁹</td>
<td>eMurmur ID</td>
<td>CSD Labs</td>
<td>PMN</td>
<td>April 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automated analysis of specific heart sounds, including S1, S2, physiological and pathological heart murmurs.</td>
<td>Phonocardiogram</td>
<td>Case level finding of disease</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Heart murmur (i)</td>
</tr>
<tr>
<td>K192004⁵⁹</td>
<td>Eko Analysis Software</td>
<td>Eko Devices</td>
<td>PMN</td>
<td>January 2020</td>
<td>2</td>
<td>Diagnosis</td>
<td>Detects suspected heart murmurs, atrial fibrillation, and normal sinus rhythm. Calculates heart rate, QRS duration and EMAT.</td>
<td>Phonocardiogram &amp; ECG</td>
<td>Case level finding of disease</td>
<td>NN</td>
<td>Yes</td>
<td>Assistive</td>
<td>Cardiac Arrhythmias, Heart murmur (i)</td>
</tr>
<tr>
<td>Approval no</td>
<td>Device</td>
<td>Manufacturer</td>
<td>Scheme</td>
<td>Date</td>
<td>Class</td>
<td>Clinical task</td>
<td>Device task</td>
<td>Device input</td>
<td>Device output</td>
<td>ML method</td>
<td>SaMD</td>
<td>Autonomy</td>
<td>Disease(s)</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>K192109</td>
<td>KOALA</td>
<td>IB Lab</td>
<td>PMN</td>
<td>November 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Metric measurements of the joint space width and indicators for presence of osteoarthritis on PA/AP knee X-ray images.</td>
<td>X-ray</td>
<td>Case level finding of disease</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Knee osteoarthritis (i)</td>
</tr>
<tr>
<td>DEN180005</td>
<td>OsteoDetect</td>
<td>Imagen Technologies</td>
<td>De novo</td>
<td>May 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Identify and highlight distal radius fractures during the review of posterior-anterior and lateral radiographs of adult wrists.</td>
<td>X-ray</td>
<td>Case level finding of disease</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td>Wrist fracture (i)</td>
</tr>
<tr>
<td>K180432</td>
<td>Al-ECG Platform</td>
<td>Shenzhen Carewell Electronics</td>
<td>PMN</td>
<td>November 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Interprets ECG for cardiac abnormalities, such as arrhythmias, myocardial infarction, ventricular hypertrophy, and abnormal ST-T changes.</td>
<td>ECG</td>
<td>Case level finding of disease</td>
<td>AI</td>
<td>Yes</td>
<td>Assistive</td>
<td>Cardiac arrhythmias (i)</td>
</tr>
<tr>
<td>K180125</td>
<td>Powerlook Density Assessment Software</td>
<td>iCad</td>
<td>PMN</td>
<td>April 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Analyses the dense tissue area of each breast, providing a Category of 1-4 consistent with ACR BI-RADS fifth edition.</td>
<td>Digital breast tomosynthesis</td>
<td>Clinical grading or scoring</td>
<td>ML</td>
<td>Yes</td>
<td>Decision</td>
<td>Breast cancer (a)</td>
</tr>
<tr>
<td>K172983</td>
<td>HealthCCS</td>
<td>Zebra Medical Vision</td>
<td>PMN</td>
<td>June 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Quantifies calcification in the coronary arteries reporting a risk category.</td>
<td>CT</td>
<td>Clinical grading or scoring</td>
<td>CNN</td>
<td>Yes</td>
<td>Information</td>
<td>Coronary artery disease (i)</td>
</tr>
<tr>
<td>K173542</td>
<td>Arterys Oncology DL</td>
<td>Arterys</td>
<td>PMN</td>
<td>January 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Quantification of lung nodules and liver lesions, reporting in accordance with Lung-RADS and LI-RADS.</td>
<td>CT / MRI</td>
<td>Clinical grading or scoring</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td>Lung and liver cancers. (a)</td>
</tr>
<tr>
<td>Approval no</td>
<td>Device</td>
<td>Manufacturer</td>
<td>Scheme</td>
<td>Date</td>
<td>Class</td>
<td>Clinical task</td>
<td>Device task</td>
<td>Device input</td>
<td>Device output</td>
<td>ML method</td>
<td>SaMD</td>
<td>Autonomy</td>
<td>Disease(s)</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>------------</td>
<td>-------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>K17054093</td>
<td>DM-Density</td>
<td>Densitas</td>
<td>PMN</td>
<td>February 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Analyses and reports breast density in accordance with the BI-RADS density classification scales.</td>
<td>Digital mammography</td>
<td>Clinical grading or scoring</td>
<td>AI</td>
<td>Yes</td>
<td>Assistive</td>
<td>Breast cancer (a)</td>
</tr>
<tr>
<td>K19044294</td>
<td>Koios DS for Breast</td>
<td>Koios Medical</td>
<td>PMN</td>
<td>July 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Automatically classify user-selected region(s) of interest containing a breast lesion into four BI-RADS-aligned categories.</td>
<td>Ultrasound</td>
<td>Clinical grading or scoring</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Breast cancer (i)</td>
</tr>
<tr>
<td>K17056894</td>
<td>Cardiologs</td>
<td>Cardiologs Technologies</td>
<td>PMN</td>
<td>June 2017</td>
<td>2</td>
<td>Diagnosis</td>
<td>Identification and labelling abnormal cardiac rhythms.</td>
<td>ECG</td>
<td>Identify features of disease</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td>Cardiac Arrhythmias (i)</td>
</tr>
<tr>
<td>P0800003/S00895</td>
<td>Selenia Dimensions 3D System High Resolution Tomosynthesis</td>
<td>Hologic</td>
<td>PMA</td>
<td>October 2019</td>
<td>3</td>
<td>Diagnosis</td>
<td>Identification of clinically relevant regions of interest for breast screening and diagnosis.</td>
<td>Digital breast tomosynthesis</td>
<td>Identify features of disease</td>
<td>AI</td>
<td>No</td>
<td>Assistive</td>
<td>Breast cancer (i)</td>
</tr>
<tr>
<td>K19199496</td>
<td>Profound AI</td>
<td>iCad</td>
<td>PMN</td>
<td>October 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Detect malignant soft tissue densities and calcifications in breast images.</td>
<td>Digital breast tomosynthesis</td>
<td>Identify features of disease</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td>Breast cancer (a)</td>
</tr>
<tr>
<td>DEN17002297</td>
<td>Quantx</td>
<td>Quantitative Insights</td>
<td>De novo</td>
<td>July 2017</td>
<td>2</td>
<td>Diagnosis</td>
<td>Analysis of selected regions, providing comparison with similar lesions with a known ground truth.</td>
<td>MRI</td>
<td>Identify features of disease</td>
<td>AI</td>
<td>Yes</td>
<td>Assistive</td>
<td>Breast cancer (a)</td>
</tr>
<tr>
<td>K16120198</td>
<td>ClearRead CT</td>
<td>Riverain Technologies</td>
<td>PMN</td>
<td>September 2016</td>
<td>2</td>
<td>Diagnosis</td>
<td>Generates secondary vessel suppressed lung images, marking regions of interest to aid in the detection of pulmonary nodules.</td>
<td>CT</td>
<td>Identify features of disease</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Lung cancers (a)</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Approval no</th>
<th>Device</th>
<th>Manufacturer</th>
<th>Scheme</th>
<th>Date</th>
<th>Class</th>
<th>Clinical task</th>
<th>Device task</th>
<th>Device input</th>
<th>Device output</th>
<th>ML method</th>
<th>SaMD</th>
<th>Autonomy</th>
<th>Disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K192287</td>
<td>Transpara</td>
<td>ScreenPoint Medical BV</td>
<td>PMN</td>
<td>December 2019</td>
<td>2</td>
<td>Diagnosis</td>
<td>Identify regions suspicious for breast cancer and assess their likelihood of malignancy.</td>
<td>Digital mammography</td>
<td>Identify features of disease</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Breast cancer (i)</td>
</tr>
<tr>
<td>K182218</td>
<td>FerriSmart Analysis System</td>
<td>Resonance Health Analysis Services</td>
<td>PMN</td>
<td>November 2018</td>
<td>2</td>
<td>Diagnosis</td>
<td>Measure liver iron concentration.</td>
<td>MRI</td>
<td>Quantification</td>
<td>CNN</td>
<td>Yes</td>
<td>Decision</td>
<td></td>
</tr>
<tr>
<td>K191370</td>
<td>DreaMed Advisor Pro</td>
<td>DreaMed Diabetes</td>
<td>PMN</td>
<td>July 2019</td>
<td>2</td>
<td>Treatment</td>
<td>Analyse blood glucose and insulin pump data to generate recommendations for optimising insulin pump dose ratios.</td>
<td>Blood glucose and insulin pump data</td>
<td>Treatment recommendations</td>
<td>ML</td>
<td>Yes</td>
<td>Assistive</td>
<td>Type 1 diabetes (i)</td>
</tr>
<tr>
<td>K190896</td>
<td>Briefcase (CSF)</td>
<td>AiDoc Medical</td>
<td>PMN</td>
<td>May 2019</td>
<td>2</td>
<td>Triage</td>
<td>Analyse images, notifying cases with suspected positive findings of linear lucencies in the cervical spine bone in patterns compatible with fractures.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Cervical spine fracture (i)</td>
</tr>
<tr>
<td>K190072</td>
<td>Briefcase (ICH and PE)</td>
<td>AiDoc Medical</td>
<td>PMN</td>
<td>April 2019</td>
<td>2</td>
<td>Triage</td>
<td>Analyse images, notifying cases with suspected positive findings of Intracranial Haemorrhage or Pulmonary Embolism pathologies.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Intracranial Haemorrhage, Pulmonary Embolism (i)</td>
</tr>
<tr>
<td>K192383</td>
<td>Briefcase (LVO)</td>
<td>AiDoc Medical</td>
<td>PMN</td>
<td>December 2019</td>
<td>2</td>
<td>Triage</td>
<td>Analyse images, notifying cases with suspected positive findings of Large Vessel Occlusion pathologies.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Large Vessel Occlusion (i)</td>
</tr>
<tr>
<td>K183285</td>
<td>CmTriage</td>
<td>CureMetrix</td>
<td>PMN</td>
<td>March 2019</td>
<td>2</td>
<td>Triage</td>
<td>Analyse screening mammograms, notifying cases with at least one suspicious finding at the exam level.</td>
<td>Digital mammography</td>
<td>Triage notifications</td>
<td>ML</td>
<td>Yes</td>
<td>Decision</td>
<td>Breast cancer (a)</td>
</tr>
<tr>
<td>Approval no</td>
<td>Device</td>
<td>Manufacturer</td>
<td>Scheme</td>
<td>Date</td>
<td>Class</td>
<td>Clinical task</td>
<td>Device task</td>
<td>Device input</td>
<td>Device output</td>
<td>ML method</td>
<td>SaMD</td>
<td>Autonomy</td>
<td>Disease(s)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>------------</td>
<td>-------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>K183182</td>
<td>Critical Care Suite</td>
<td>GE Medical Systems</td>
<td>PMN</td>
<td>August 2019</td>
<td>2</td>
<td>Triage</td>
<td>Analyse and notify cases with suspected findings of pneumothorax.</td>
<td>X-ray</td>
<td>Triage notifications</td>
<td>DL</td>
<td>Yes</td>
<td>Decision</td>
<td>Pneumothorax (i)</td>
</tr>
<tr>
<td>K182177</td>
<td>Accipiolx</td>
<td>MaxQ-AI</td>
<td>PMN</td>
<td>October 2018</td>
<td>2</td>
<td>Triage</td>
<td>Identify and notify cases with suspected findings of acute intracranial haemorrhage.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Intracranial Haemorrhage (i)</td>
</tr>
<tr>
<td>DEN170073</td>
<td>ContaCT</td>
<td>Viz AI</td>
<td>De novo</td>
<td>February 2018</td>
<td>2</td>
<td>Triage</td>
<td>Identify and notify cases with suspected findings of large vessel occlusion.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Large Vessel Occlusion (i)</td>
</tr>
<tr>
<td>K192320</td>
<td>HealthCXR</td>
<td>Zebra Medical Vision</td>
<td>PMN</td>
<td>November 2019</td>
<td>2</td>
<td>Triage</td>
<td>Automatically identify and notify suspected findings of pleural effusion.</td>
<td>X-ray</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Pleural Effusion (i)</td>
</tr>
<tr>
<td>K190424</td>
<td>HealthICH</td>
<td>Zebra Medical Vision</td>
<td>PMN</td>
<td>June 2019</td>
<td>2</td>
<td>Triage</td>
<td>Automatically identify and notify suspected findings of Intracranial Haemorrhage.</td>
<td>CT</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Intracranial Haemorrhage (i)</td>
</tr>
<tr>
<td>K190362</td>
<td>HealthPNX</td>
<td>Zebra Medical Vision</td>
<td>PMN</td>
<td>May 2019</td>
<td>2</td>
<td>Triage</td>
<td>Automatically identify and notify suspected findings of Pneumothorax.</td>
<td>X-ray</td>
<td>Triage notifications</td>
<td>AI</td>
<td>Yes</td>
<td>Decision</td>
<td>Pneumothorax (i)</td>
</tr>
</tbody>
</table>

**Action Implementation**

<table>
<thead>
<tr>
<th>Approval no</th>
<th>Device</th>
<th>Manufacturer</th>
<th>Scheme</th>
<th>Date</th>
<th>Class</th>
<th>Clinical task</th>
<th>Device task</th>
<th>Device input</th>
<th>Device output</th>
<th>ML method</th>
<th>SaMD</th>
<th>Autonomy</th>
<th>Disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K191713</td>
<td>FluoroShield</td>
<td>Omega Medical Imaging</td>
<td>PMN</td>
<td>October 2019</td>
<td>2</td>
<td>Procedure</td>
<td>Automated region of interest in fluoroscopic applications.</td>
<td>Fluoroscopy</td>
<td>Automatic control of device</td>
<td>AI</td>
<td>No</td>
<td>Decision</td>
<td></td>
</tr>
<tr>
<td>DEN190040</td>
<td>Caption Guidance</td>
<td>Bay Labs</td>
<td>De novo</td>
<td>February 2020</td>
<td>2</td>
<td>Procedure</td>
<td>Real-time guidance during echocardiography acquisition to assist obtaining correct images of standard diagnostic views and orientations.</td>
<td>Echocardiogram</td>
<td>Automatic control of device</td>
<td>DL</td>
<td>Yes</td>
<td>Assistive</td>
<td></td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Indicated or associated disease*</th>
<th>N</th>
<th>Diagnose</th>
<th>Triage</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intracranial haemorrhage</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Lung cancer</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Heart murmurs</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Large vessel occlusion</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical spine fracture</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Liver cancer</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sleep and respiratory related sleep disorders</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus type 1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Wrist fracture</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

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

Table 3  ML device output by type

<table>
<thead>
<tr>
<th>Output type</th>
<th>Devices</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantification</td>
<td>13</td>
<td>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.54</td>
</tr>
<tr>
<td>Triage notifications</td>
<td>10</td>
<td>Triage notification alert clinicians to cases with suspected positive findings.25 40–48</td>
</tr>
<tr>
<td>Case-level finding of disease</td>
<td>6</td>
<td>Case level finding of disease such as, wrist fractures,52 diabetic retinopathy,49 osteoarthritis,51 heart murmurs24 50 and cardiac arrhythmias.24 53</td>
</tr>
<tr>
<td>Identify features of disease</td>
<td>6</td>
<td>Identify features of disease thereby drawing clinician attention to them, such prompting breast55–57 59 or lung58 cancers on images or cardiac arrhythmias on ECG tracings.54</td>
</tr>
<tr>
<td>Clinical grading or scoring</td>
<td>5</td>
<td>Clinical grading or scoring (n=5) on standardised clinical assessment instruments, such as BI-RADS,60 63 64 LI-RADS,65 lung-RADS,65 or Agatston-equivalent scores.61</td>
</tr>
<tr>
<td>Enhanced images</td>
<td>4</td>
<td>Enhanced images with reduced noise and improved image quality.26–29</td>
</tr>
<tr>
<td>Automatic coding of features or events</td>
<td>2</td>
<td>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.58</td>
</tr>
<tr>
<td>Automatic control of electronic or mechanical devices</td>
<td>2</td>
<td>Automatic control of electronic or mechanical devices, such as fluoroscopy collimator67 and automatic recording of ultrasound clips dependent on detected image quality.66</td>
</tr>
<tr>
<td>Treatment recommendations</td>
<td>1</td>
<td>Treatment recommendations, such as adjustments to insulin pump dose ratios.87</td>
</tr>
</tbody>
</table>
Profound AI, 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, which reports breast density grading for digital monography cases according to the American Collage of Radiology’s Breast Imaging-Reporting and Data System Atlas.

Action implementation
Devices providing action implementation included Caption Guidance and FluroShield; these implemented decisions through the automatic control of an electronic or mechanical device. Caption Guidance 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. FluroShield 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, five of six devices reporting a case level finding of disease, and almost half of devices providing quantification. 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, but also included IDx-DR 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, quantification, and one device which coded features or events.

**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.'
- 'Not intended for making clinical decisions regarding patient treatment or for diagnostic purposes.'
- 'Intended as an additional input to standard diagnostic pathways and is only to be used by qualified clinicians.'
- 'Interpretations offered by (device) are only significant when considered in conjunction with healthcare provider over-read and including all other relevant patient data.'
- 'Should not be used in lieu of full patient evaluation or solely relied on to make or confirm a diagnosis.'
- 'The clinician retains the ultimate responsibility for making the pertinent diagnosis based on their standard practice.'
- 'Patient management decisions should not be made solely on the results.'
- 'Provides adjunctive information and is not intended to be used without the original CT series.'

**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 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. 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, identifies risks associated with false-negatives that could
lead to incorrect or delayed patient management, while false-positives may deprioritise other cases.

Likewise, the diabetic retinopathy screening device, IDx-DR\textsuperscript{46} 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{9, 73} Troublingly, people who suffer these automation biases exhibit reduced information seeking\textsuperscript{74–76} and reduced allocation of cognitive resources to process that information,\textsuperscript{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,\textsuperscript{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.\textsuperscript{78, 79} Greater consequences could be due to reduced situational awareness as automation takes over more stages of human information processing.\textsuperscript{79}

Indeed, the De Novo for Quantx,\textsuperscript{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 OsteoDetect\textsuperscript{52} 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.’\textsuperscript{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.\textsuperscript{9, 60}

Other risks identified in De Novo classifications\textsuperscript{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 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,\textsuperscript{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.\textsuperscript{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.\textsuperscript{12} Presently, the FDA only approves devices with ‘locked’ algorithms,\textsuperscript{82} but are moving towards a framework that would permit ML devices which learn and adapt to real-world data.\textsuperscript{83} Such a framework is expected to involve precertification of vendors and submission

\textsuperscript{49}Open access


of algorithm change protocols.\textsuperscript{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 techniques\textsuperscript{84} 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.\textsuperscript{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 careful 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 @farrahmagrabi

**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**


81 International Medical Device Regulators Forum. “Software as a Medical Device”: Possible Framework for Risk Categorization and Corresponding Considerations, 2014.


83 U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, 2021.


APPENDIX A: SEARCH STRATEGY AND SELECTION METHOD

It is not possible to identify ML devices directly from FDA databases, which function as registries of approvals.[1-3] While searchable by fields such as device name, manufacturer and product classification, there is no capability to search by the content of approvals or the algorithmic methods used by devices. Therefore, we conducted a systematic internet search to identify candidate ML devices. We then searched the FDA databases for those candidate devices and retrieved all publicly available approval documents. Each candidate was then assessed for inclusion using the approval documents.

Inclusion criteria

Medical devices that were:

1. FDA approved, evidenced by the existence of a premarket approval (PMA), premarket notification (PMN), or a De Novo classification.
2. Intended for use by clinicians or healthcare professionals.
3. Intended to be used for or assist with clinical tasks/decisions.
4. Utilised ML.

Search strategy

An internet search (www.google.com) for candidate ML devices using query combinations of “machine learning” and “FDA approved” (see Box 1) was performed in February 2020 and the resulting URLs were captured. The search was conducted using Firefox 72.0.2 (64-bit) by Mozilla Foundation utilising private browsing mode and not signed into Google or any other account that may impact search results. A virtual Private Network (VPN) allowed to search to be conducted from within the USA, that is, within the FDA’s jurisdiction.

URLs were accessed in order and any candidate devices mentioned were extracted. Extraction ceased after the 50th result having reached saturation, with the last 18 results not yielding any new candidates. We also extracted candidates from a list of ‘FDA Cleared AI Algorithms’ curated by the American College of Radiology’s Data Science Institute [4] supplemented by additional hand searching.

Selection of ML devices

Two reviewers (D.L. and P.S) independently assessed the candidate devices against the inclusion criteria. For ML utilisation, device manufacturer marketing materials and the product website were also examined. Any claim indicative of utilisation of machine-learned algorithms, including specific mention or general descriptions of ML techniques or methods were considered.

Interrater agreement was fair to moderate ($\kappa = .44$, 95%CI .32 to .57).[5] Disagreements were independently reviewed by a third reviewer (J.C.) and resolved by consensus by all three reviewers. Seven devices had multiple approvals, with latter approvals modifying the device in a way that required a new approval. Therefore, the most recent approval for each unique device was considered. In total, our 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.

Box 1 Search query

("artificial intelligence" OR "ai" OR "machine learning" OR "deep learning") AND ("FDA approved" OR "FDA approves" OR "FDA approval")
Box 2 Examples of FDA approved indications specifying responsibility for the final decision on the device task resides with the clinician.

“All automatically scored events are subject to verification by a qualified clinician.” [6]

Device “is a support tool that provides relevant clinical data as a resource to the clinician and is not intended to be a source of medical advice or to determine or recommend a course of action or treatment for a patient.” [7]

“Not intended for making clinical decisions regarding patient treatment or for diagnostic purposes.” [8]

“not intended to diagnose, treat, or prevent diseases or conditions” [9]

“intended to provide the trained medical professional with complementary information for the evaluation and assessment of MR brain images and to aid the trained medical professional in quantitative reporting.” [10 11]

“intended as an additional input to standard diagnostic pathways and is only to be used by qualified clinicians.” [12]

“interpretations offered by [device] are only significant when considered in conjunction with healthcare provider over-read and including all other relevant patient data.” [13]

“not intended as a sole means of diagnosis.” [14]

Device “should not be used in-lieu of full patient evaluation or solely relied upon to make or confirm a diagnosis.” [15]

“Clinicians should review [device] annotated images concurrently with original images before making a final determination on a case. [device] is an adjunct tool and does not replace the role of the clinician. Clinicians must not use the CAD generated output as the primary interpretation.” [16]

“interpretation results are not intended to be the sole means of diagnosis for any abnormal ECG. They are offered to physicians and clinicians on an advisory basis only in conjunction with the physician’s knowledge of ECG.” [17]

“The clinician retains the ultimate responsibility for making the pertinent diagnosis based on their standard practices.” [18]

“to aid interpreting physicians in the assessment of breast tissue composition. [device] produces adjunctive information. It is not a diagnostic aid.” [19]

“Patient management decisions should not be made solely on the results of the” device. [20]

“not intended to be the sole means of diagnosis. It is offered to physicians and clinicians on an advisory basis only in conjunction with the physician’s knowledge of ECG patterns, patient background, clinical history, symptoms, and other diagnostic information.” [21]

Device output “may be confirmed or dismissed by the interpreting physician.” [22]

“not intended for primary interpretation of digital mammography images.” [23]

“provides adjunctive information and is not intended to be used without the original CT series.” [24]

“Patient management decisions should not be made solely on the basis of analysis by” device. [25]

Device “does not replace clinical judgement.” [26]

“Product users are responsible for image quality and diagnosis.” [27]
LIST OF REFERENCES


22. U.S. Food and Drug Administration. K191994, ProFound AI Software V2.1, iCAD Inc. 510(k) Premarket Notification; 2019

23. U.S. Food and Drug Administration. DEN170022, QuantX, Quantitative Insights, Inc. Device Classification Under Section 513(f)(2)(De Novo); 2017


27. U.S. Food and Drug Administration. DEN190040, Caption Guidance, Bay Labs, Inc. Device Classification Under Section 513(f)(2)(De Novo); 2020