Clinical Radiology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Review

Artificial intelligence in breast imaging

E.P.V. Le [a,b], Y. Wang [b], Y. Huang [b,c], S. Hickman [c], F.J. Gilbert [b,c,*]

[a] University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK

[b] EPSRC Centre for Mathematical and Statistical Analysis of Multimodal Clinical Imaging, University of Cambridge, Cambridge CB3 0WA, UK

[c] Department of Radiology, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK

This article reviews current limitations and future opportunities for the application of computer-aided detection (CAD) systems and artificial intelligence in breast imaging. Traditional CAD systems in mammography screening have followed a rules-based approach, incorporating domain knowledge into hand-crafted features before using classical machine learning techniques as a classifier. The first commercial CAD system, ImageChecker M1000, relies on computer vision techniques for pattern recognition. Unfortunately, CAD systems have been shown to adversely affect some radiologists' performance and increase recall rates. The Digital Mammography DREAM Challenge was a multidisciplinary collaboration that provided 640,000 mammography images for teams to help decrease false-positive rates in breast cancer screening. Winning solutions leveraged deep learning's (DL) automatic hierarchical feature learning capabilities and used convolutional neural networks. Start-ups Therapixel and Kheiron Medical Technologies are using DL for breast cancer screening. With increasing use of digital breast tomosynthesis, specific artificial intelligence (AI)-CAD systems are emerging to include iCAD's PowerLook Tomo Detection and ScreenPoint Medical's Transpara. Other AI-CAD systems are focusing on breast diagnostic techniques such as ultrasound and magnetic resonance imaging (MRI). There is a gap in the market for contrast-enhanced spectral mammography AI-CAD tools. Clinical implementation of AI-CAD tools requires testing in scenarios mimicking real life to prove its usefulness in the clinical environment. This requires a large and representative dataset for testing and assessment of the reader's interaction with the tools. A cost-effectiveness assessment should be undertaken, with a large feasibility study carried out to ensure there are no unintended consequences. AI-CAD systems should incorporate explainable AI in accordance with the European Union General Data Protection Regulation (GDPR).

Introduction

Globally, breast cancer is the most common cancer amongst women[1] and remains the second leading cause of cancer death.[2] The importance of mammography screening

* Guarantor and correspondent: F. J. Gilbert, Department of Radiology, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK. Tel.: +44 1223 746439.
E-mail address: fjg28@cam.ac.uk (F.J. Gilbert).

https://doi.org/10.1016/j.crad.2019.02.006
0009-9260/

has been widely recognised[3] and the early detection of breast cancer is associated with better outcomes.[4,5]

The limitations of mammography screening include over-diagnosis,[6] overtreatment and false-positive rates with associated negative psychological impact[7,8] and unnecessary costs and biopsies.[3] Interpretation of mammograms varies with experience,[9] is subjective,[10] and prone to error due to the heterogeneous presentation of breast cancer and the masking effect with dense breast tissue. This contributes to interval cancers that are diagnosed between screening or cancers found in the subsequent screening round.[11,12] Although reading time for two-dimensional (2D) mammography is <30–60 seconds, the huge volumes of mammograms and double-reading of each mammogram creates manpower problems and resource issues.

Computer-aided detection (CAD) systems for mammography have been used extensively in the USA following initial promising results and reimbursement introduced in 2001; however, more recently, CAD has been shown to adversely affect some radiologists' performance, increase recalls without improvement in cancer detection rates.[13] This has resulted in some scepticism as to whether or not artificial intelligence (AI) tools can reliably assist radiologists in breast cancer screening. This article reviews the current limitations and future opportunities for the application of CAD systems and artificial intelligence in breast imaging.

Brief overview of AI

Machine learning (ML) refers to the training of a model that learns features and associated parameters that most likely describe the observed data.[14] Deep learning (DL) is one such ML technique that builds upon deep neural networks, which mimic but oversimplify human neurons in the brain. DL relies on deep layered architecture that enables hierarchical learning, extracting features gradually from simple to complex abstractions from the data.[15] ML has a data-driven approach learning a mathematical model based on the observed "training" data. The goal is to make later predictions for new "test" data based on the learnt model. "Learning" or "training" in a ML regime usually refers to the iterative procedure of measuring the gap between current predictions (e.g., a cancerous mass) against a benign mass or normal tissue using an evaluation metric known as the "objective function" (also known as the "cost" or "loss" function). The ultimate aim is that with further training, the model adapts to reduce this gap in a clinically useful manner, i.e., until the predicted label matches the ground truth.

Supervised learning[16] refers to the training procedure where the observed training data and the associated ground truth labels for that data (or sometimes referred to as "targets") are both required for training the model. For instance, in mammography, a cancer is accurately outlined on the image (as the labels) allowing the algorithm to "learn" the features of malignancy from the labelled annotations. In contrast, unsupervised learning[17] is where the training data has no diagnosis or normal/abnormal labels. Semi-supervised learning or weakly supervised[18] learning provides some information to the algorithm, but not necessarily for all the training examples. Currently, supervised learning seems to be the most popular approach in image classification tasks.

Classical ML techniques such as support vector machines (SVM),[19] decision trees (DT),[20] random forests (RF),[21] and dimensionality reduction techniques, such as principal component analysis (PCA), have relatively low computational requirements, as these models involve relatively fewer parameters compared to DL algorithms. These algorithms are actively used in computer vision for feature extraction and feature selection of hand-crafted features. DL is more time-consuming in training the models as DL tends to have millions of parameters and hence requires high-performance computing hardware.

There are various DL architectures published in the literature, but the majority of these networks fundamentally build upon some basic and similar neural network building blocks, referred to as "layers".[22] A neural network consists of consecutive layers, comprising an input layer (e.g., raw pixels of a mammogram), hidden layer(s), and an output layer (e.g., outputs the prediction of the model: "benign"/"malignant" label). This is shown in Fig 1. The earlier layers in a DL model are considered to act similarly to the simple cells of the human primary visual cortex that learn low-level features such as edges in particular orientations in the image. Higher levels of abstraction are a result of multiple stacked layers.[23] Information is propagated through the DL architecture, and more complex features are extracted. These features are finally fed through the last layer of the network architecture for prediction or classification.[15]

Convolutional neural networks (CNNs)[15,24] consist of special ML structures that are one of the most popular architectures in general image analysis applications today. CNNs have characteristic layers in which a convolution operation (also referred to as a kernel) acts as a filter on the pixel image matrix to extract spatially correlated features of

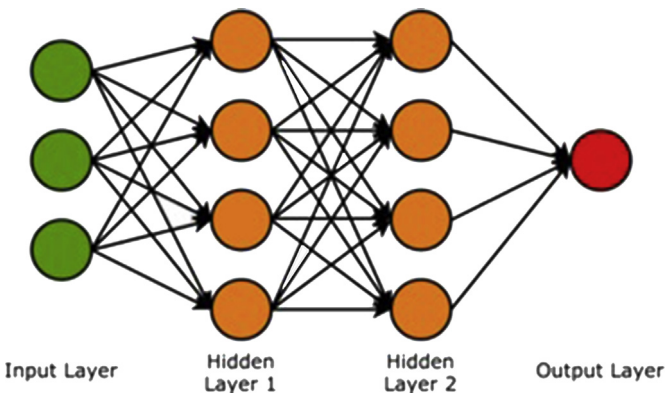

Figure 1 Structure of a neural network. Groups of neurons that perform similar functions are aggregated into layers. A neural network consists of consecutive layers, comprising an input layer, a hidden layer(s), and an output layer.

Please cite this article as: Le EPV et al., Artificial intelligence in breast imaging, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.02.006

the input image and hence offer some shift invariance. The convolutional layer generates a linear output (feature map). Other layers include the pooling and fully connected layers.[23,25] The feature map created in the convolutional layer is initially passed through a non-linear activation function, e.g., a rectified linear unit (ReLU) that converts all negative values to 0, mapping it to an output. This is then transferred to the pooling layer to enable down-sampling of the feature map. Lastly, the output is then passed into the fully connected layer to classify the overall outcome. These layers constitute components of the hidden layers of a CNN. There can be hundreds of hidden layers and with an increased number of layers, there is an increase in the complexity of feature extraction.

Training a CNN requires a vast amount of data, significantly more than required for other types of ML. Radiology is one of the most digitised fields in medicine; however, the majority of the data is not freely accessible, not labelled appropriately, or of the quality required. To overcome this limitation in training, transfer learning (TL)[26] can be applied. This works by pre-training a model to detect various key features, e.g., edges and then apply this pre-existing learning to a related image task, such as identification of lesions on a mammogram. Subsequently processes of feature extraction or fine tuning can occur to adjust the network for the new feature detection task.[25] Other methods to overcome limitations of the available data include, data augmentation techniques to minimise overfitting, unsupervised[17] learning or utilisation of generative adversarial networks (GANs).[25]

CAD systems in mammography screening

The first commercial CAD system, ImageChecker M1000 system (R2 Technology, Los Altos, CA, USA), received US Food and Drug Administration (FDA) approval in 1998 with the aim of providing a prompt or second opinion to the radiologist in order to enhance radiologists' diagnostic accuracy. ImageChecker is a computer-assisted detection system rather than a standalone diagnostic tool.

The R2 approach utilised classical computer vision techniques for pattern recognition. It searched for clusters of bright spots, which are suggestive of microcalcifications, and radial lines suggestive of spiculate masses within a concentric "annulus" between 3 and 16 mm radii. The R2 system calculated the likelihood of malignant microcalcifications and malignant masses, and highlighted regions above a fixed threshold of likelihood for radiologists.[27]

Traditional CAD schemes followed a similar rules-based approach,[28,29] incorporating prior domain knowledge into hand-crafted features before using classical ML techniques as a classifier.[30] For example, the earliest breast cancer detection CAD approach published in the academic literature incorporated the importance of global side-to-side comparison for each breast to identify any areas of architectural distortion or asymmetry, which could indicate malignancy.[31]

Mathematically describing qualitative visual characteristics and patterns of cancer is time-consuming and prone to subjectivity: the method that relevant features are chosen (known as feature engineering) and how these features are combined leads to different outputs once they are fed into a ML classifier. Furthermore, there is no consensus on the most suitable ML technique for classification in the final stage of the CAD pipeline. These ML techniques aim to classify breast tissue lesions based on the inputted extracted features by identifying hidden patterns and non-linear relationships between these variables. To date the SVM[19,32] classifier has been the most widely used technique in traditional CAD systems.[33]

The accuracy of these classification algorithms depends heavily on the prior feature engineering process and is limited by the available expert knowledge and the ability to hand-engineer suitable descriptive features. It is difficult to evaluate the performance of products from different research groups as they are tested on various private datasets or public domain datasets[34–36] without standardisation.

In the UK screening programme, a single reader using traditional CAD was compared with standard double reading. Both methods had similar cancer detection rates, but the recall rates for single reading with CAD increased by 0.5% compared to double reading.[37] Although CAD systems have been around for many decades, they have yet to reach satisfactory clinical utility.[13,38] It has been difficult to translate radiologist intuition using the traditional ML approach. CAD systems face additional challenges related to the low signal-to-noise ratio in mammographic images as well as the large variation in lesion appearance: location, shape, size, and boundaries.

Application of ML to the mammographic screening population could provide automated assistance in detection of breast cancer. Testing artificial neural networks (ANNs) against three readers on 18 cancers and 233 controls produced a sensitivity of 73.7% versus 66.6% and specificity of 72% versus 92.7%.[39] Another approach to detect malignancy and rule out normal mammograms found a sensitivity of 76.1% with a specificity of 88.5% for both screening and diagnostic mammograms from three mammography vendors.[40] This study used a deep CNN without pixel level supervision and was trained by weak supervision. In testing, the specificity decreased from 100% to 77.78% when density increased.

Applying pre-trained CNNs to distinguish benign from malignant lesions achieved an area under the receiver operating characteristic curve (AUC) of 0.99 when tested on an independent dataset of 113 mammograms with a radiologist-defined ground truth.[41] A further study applied the publicly available INbreast dataset[35] to test classification of benign and malignant cases and achieved an AUC of 0.91 ($\pm$0.12) versus 0.76 ($\pm$0.23) for a manual assessment.[42]

AI application for breast density

Several software programs receiving FDA clearance have automated breast density assessment as higher breast

density confers a greater risk of breast cancer[43] with high interobserver variability in breast density estimation.[44] Masking can occur in dense breast tissue preventing detection of cancers.[45] An automated assessment of breast density with consistent quantitative classification could guide the need for supplemental imaging with ultrasound, contrast-enhanced mammography, or magnetic resonance imaging (MRI).

Research into the classification of Breast Imaging-Reporting and Data System (BI-RADS) categories is limited by the variable ground truth provided by human readers.[46] A clinical implementation study demonstrated 77% agreement between a single radiologist's BI-RADS classification and a deep CNN algorithm when tested on 8,677 previously unseen mammograms, as well as overall acceptability into daily workflow.[47] The highest rates of disagreement occurred between scattered fibroglandular tissue and heterogeneously dense category (BIRADS B and C). Seeking consistent classification between breast density categories (B and C) a CNN was tested on 1,850 processed non-cancer images acquired from a single vendor, with quantitative selection of cases to ensure correctly labelled data were used. An AUC of 0.9882 versus 0.9857 was achieved for an algorithm without pre-training and with pre-training.[48]

VolparaDensity (Volpara Solutions) derives the thickness of fibroglandular tissue and the thickness of fat from areas of brightness in the mammography image, which can be collectively used to calculate the volumetric breast density.[49] Hologic received FDA 510(k) clearance in December 2017 for their Quantra 2.2 Breast Density Assessment software,[50] which uses ML to analyse the mammographic breast tissue distribution and texture in order to output one of four categories in alignment with the US breast density rating system: BI-RADS 5th edition.[51] Densitas received FDA 510(k) clearance in February 2018[52] for their ML-based automated breast density software, which calculates percent breast density and outputs an automated report with breast density grade and BI-RADS density category.

The DM DREAM challenge and DL in mammography

The Digital Mammography DREAM Challenge (DM DREAM),[53] which ran from November 2016 to May 2017, was a multidisciplinary collaboration[54]: Sage bionetworks provided the infrastructure for the competition, the Group Health Cooperative (a National Cancer Institute-funded Breast Cancer Surveillance Consortium registry) contributed more than 640,000 de-identified digital mammography images, and teams from all over the world submitted their solutions to improve diagnostic accuracy in breast cancer screening. The top winning teams incorporated different aspects of ML with a strong focus on DL. DL approaches include the ability to deal with an increased complexity of input data, where raw imaging data rather than statistically derived features, such as entropy and kurtosis,[55] are used. Each pixel in a medical image represents one dimension, therefore medical images are extremely high-dimensional that are computationally expensive. Secondly, compared with classical ML approaches, the models are able to learn much more interesting non-linear relationships between the input and output.

Despite its vast potential, DL is not a magic bullet and mammography presents many challenges that go beyond fine-tuning the parameters of a DL model. As illustrated by one of the winning teams of the DM DREAM competition, challenges include (1) dealing with an imbalanced dataset, (2) processing high-resolution images, and (3) weakly supervised learning.

Only 0.35% (1,114 images out of 318,000) of the training data for the DM DREAM challenge were breast cancers.[56] The first sub-challenge of the DM DREAM challenge was to predict the cancer status of each breast of a patient, outputting a score between 0 and 1, using only their screening mammogram. This competition focused on a combination of both computer-aided detection as well as computer-assisted diagnosis. Unlike other public domain mammography datasets, such as the commonly used Digital Database for Screening Mammography (DDSM)[35] dataset (which has pixel-level annotation), the DM DREAM training data did not include cancerous lesion location. Instead this was an example of weakly supervised learning as only mammogram-level labels, 0 or 1, to indicate absence or presence of cancer, were provided. The ML algorithms had to learn discriminatory features between cancer and non-cancer cases from very few positive examples.

Mammography images have a large memory requirement as they are an order of magnitude larger than the commonly used natural images in ML and computer vision tasks ($\sim$3,000$\times$3,000 pixels in mammography versus $\sim$300$\times$300 pixels in ImageNet[57] competition). Significantly down-scaling a whole mammographic image, down-sampling with excessive pooling or focusing on classifying only a small region of interest to reduce computational cost would hamper performance as breast cancer screening relies heavily on the fine details. Microcalcifications and other cancerous lesions are relatively small and subtle. In addition, whereas high resolution enables in-depth examination of the finer details, it is also important to visualise the whole breast in order to assess for global features such as architectural distortion. The ML approach to mammography data must not only deal with a single high-resolution image, but should have the ability to take into consideration the standard four views concurrently and ideally prior mammograms as sometimes cancers are only visible in a certain view, as well as allowing for evaluation of breast asymmetries.[58]

Therapixel, a French start-up was one of the overall top teams in the DM DREAM challenge, provided a solution by using a multi-stage approach and CNNs.[56] They started by pre-training a lesion detector (called Detector Net) using the external Digital Database for Screening Mammography (DDSM) dataset[35] and TL (employing a modified VGG architecture), followed by end-to-end fine-tuning on the whole mammography images themselves for binary classification: label 0 = normal and label 1 = cancer. TL[59] refers

Please cite this article as: Le EPV et al., Artificial intelligence in breast imaging, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.02.006

to applying a model trained on a different dataset (such as natural images) to a different domain (medical images). Lower-level features are learnt that are useful in both natural images and medical image applications; however, there are fundamental differences between the low-resolution, 2D, colour (three-channel) images, i.e., the natural images, in ImageNet[60] and medical images such as mammograms and this shift in domain between natural images and medical images impinges on the performance of the TL method. Therapixel's best-performing model achieved an AUC of 0.85, a good result given the limited time available in the competitive phase and limited direct access to the DM DREAM challenge training data. The participants of the challenge could only access 500 mammography images from the DM DREAM challenge dataset directly. For the actual training of their models, the participants had to transport their models in Docker containers to the cloud for training on the large dataset in an effort to maintain suitable data protection.

In the final task both the Ann Arbour and Therapixel groups applied DL approaches to finish in joint first place.[61] The DM DREAM challenge has now entered the collaborative community phase whereby top performing teams will work together to improve their algorithms and share their source code publicly for academic and commercialisation efforts.[52,61] Collaboration sits at the heart of spurring on development for AI-CAD tools. Countries with a national breast cancer screening program, such as the USA and UK, have a great opportunity to develop high-quality, well-annotated datasets that can help to capture the large variation in breast cancer presentation as well as normal variants. For example, the Cancer Research UK (CRUK)-funded OPTIMAM project has created an anonymised dataset of 4,307 malignant cases, 450 benign cases and 680 normal cases (as of June 2017).[62,63] The OPTIMAM group are collaborating with a number of commercial groups such as DeepMind and the UK-based start-up Kheiron Medical Technologies.[64,65]

Digital breast tomosynthesis and AI

Digital breast tomosynthesis (DBT) is a three-dimensional (3D) imaging technique that has been shown to increase breast cancer-detection rates[66] and reduce false-positive[67] results as compared with digital mammography (2D) alone.[68] The superimposition of tissues in 2D digital mammography has contributed to false positives. DBT reduces overlapping shadows and increases lesion conspicuity and reduce recall rates. Recent studies have found that DBT are particularly beneficial in the detection of masses and in women with increased breast density[69] or heterogeneously dense breasts.

A relative disadvantage is that DBT increases reading times from 50–200% due to the increased number of images.[70] There is a need for optimised CAD and diagnosis systems for DBT in order to reduce radiologist evaluation time and improve efficiency. iCAD's PowerLook Tomo Detection[71] is one of the first in this domain and received FDA clearance in 2017. The tool implements DL technology to reduce the reading time of DBT. Currently the approach involves extracting regions of interest from the 3D images and incorporating them onto a synthesised 2D image to ameliorate navigation of the tomosynthesis datasets.

With the trend of increasing use of DBT, developers of AI-CAD systems have taken this emerging imaging technique into consideration. Based in the Netherlands, ScreenPoint Medical has been equipping its DL software solution, Transpara, for multimodality applications.[72] For 2D mammography, Transpara considers multiple views to automatically identify soft-tissue and calcification lesions and finally calculate a cancer suspiciousness score (on a 10-point scale: no potential abnormalities = a low score) for each mammogram.[73] The interactive decision support nature of the software is unique compared with traditional CAD prompts, as CAD marks remain hidden unless probed by the reader.[74] An independent evaluation of Transpara version 1.3.0 applied the tool to a cancer data set of 240 mammograms from two mammography vendors, read by 14 experienced radiologists from centres in the US, and demonstrated an improved breast cancer detection without an increase in reading time[75]; however, further studies are required to demonstrate the application of this tool in a screening population with blinded readers to reduce the "laboratory effect" on these results. For DBT volumes, Transpara considers the full 3D information to quantitatively analyse soft-tissue lesions and calcifications, marking the relevant section with the hope of improving DBT reading times.[76] The Transpara Score offers potential for triaging, as mammograms can be sorted according to likelihood of detectable breast cancer.

CAD systems in other breast imaging techniques

Ultrasound

Breast ultrasound can provide additional information to further characterise mammographic findings, palpable abnormalities, and guide interventional procedures. In particular, ultrasound can increase detection of early breast cancer when used as a supplementary imaging technique in women with high breast density.[77] Automated breast ultrasound (ABUS) has been developed to overcome the limitations of inter-operator variability with handheld ultrasound (HHUS) and is able to generate ~2,000 2D ultrasound images to give a 3D representation of the breast tissue.[78]

QView Medical received FDA PMA approval for its QVCAD System (QView CAD) in 2016.[79] The CAD system for ABUS has been shown to decrease reading times (RT), with CAD-ABUS averaging an RT of 113.4 seconds per case compared to 158.3 seconds per case using ABUS alone ($p<0.001$)[80] without compromising diagnostic accuracy. Development of the QVCAD DL algorithms involved global collection of ABUS cancer cases, benign and normal cases (over one million ABUS 3D images).[79] As a result, the QVCAD System is able to automatically extract features

Please cite this article as: Le EPV et al., Artificial intelligence in breast imaging, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.02.006

from suspicious areas of breast tissue with a diameter of 5 mm or more to finally generate a score of suspiciousness for each area.[81] The CAD output includes the CAD Navigator image, which is displayed simultaneously with the original ABUS and acts as a roadmap for navigation as well as CAD marks (coloured circles) for potentially malignant lesions.

Koios Medical (formerly Clearview Diagnostics) provides a cloud-based AI decision-support software for diagnostic breast ultrasound[82] called Koios DS. Following integration of the software with PACS, a physician simply clicks on a suspicious area and Koios DS can analyse over 17,000 features per region of interest (ROI) and finally provide feedback or recommendations.[83] Their algorithm was trained using a supervised ML approach that utilises thousands of radiology images and their associated pathology data.[84] This training database is constantly expanding as well, and the Koios DS system is dynamic as it is continually updated with new images and their analyses.

MRI

Breast MRI is another imaging method that may be used as an adjunct when mammography findings are inconclusive.[85] Its advantages include particularly high sensitivity and no ionising radiation. MRI forms part of the NICE guidelines for surveillance of women with a high risk of breast cancer, for example, younger women with a family history of breast cancer or those with BRCA gene mutations.[86–89]

Existing breast MRI CAD systems include CADstream (Merge Healthcare) and DynaCAD for Breast (Invivo). They are widely used in the USA[90] due to their automated breast MRI enhancements[91] via various image-processing techniques and motion correction. On the horizon, the Innovate UK-funded INTELLISCAN project focuses on bringing artificial intelligence to breast cancer MRI interpretation and reporting. This collaboration[92] involves Teeside University's Healthcare Innovation Centre (HIC), Brunel University's Innovation Centre, and First Option Software coming together to develop a software application that can provide a decision-support tool with automated anomaly detection of breast MRI examinations.[93]

Axillary response to neoadjuvant chemotherapy has previously been shown to be a strong prognostic factor. Predicting the response by use of ML could provide a further approach to utilise quantitative image data. Potential features to use as clinical markers that strongly predict the post-chemotherapy response, have been identified.[94–96] Testing of ML to predict a response in small single-institution datasets, has achieved an AUC of 0.93, using a histologically defined ground truth with T1-weighted MRI images.[97] Radiomic features allow for expansion beyond standardised radiological reporting and could provide a guide for treatments, allowing for a personalised approach to improve survival outcomes. The feasibility for ML to aid in prognostic prediction by calculation of Oncotype Dx score or predicating axillary lymph node metastasis could minimise long-term morbidities associated with axillary lymph node dissection and other invasive procedures.[98,99]

Contrast-enhanced spectral mammography (CESM)

Contrast-enhanced spectral mammography (CESM), also known as contrast-enhanced digital/dual-energy mammography (CEDM) is an enhanced form of digital mammography using iodinated contrast media. The contrast media enables better depiction of lesions against the breast background with studies suggesting that CESM has similar diagnostic performance to breast MRI, suggesting this could be an alternative option for breast MRI indications.[100,101] As this is a newly emerging technology, imaging datasets are nascent and limited and so too are CESM CAD systems. This gap in the market is a prime opportunity for AI-CAD tool developers as CESM gains growing popularity.

AI-CAD tools: limitations and challenges to implementation

Current comparison of clinical efficacy between algorithms is challenging due to variations in the datasets used and overall methodology of testing as well as a limited number of independent evaluations of efficacy.[65,102] Availability of data that are of high quality, labelled, and representative of the population, including distribution of pathology, demographics, and breast density, is limited. Current datasets commonly used in research are of small size and from single institutions and mammographic machine vendors, resulting in a potential for overfitting of algorithms. The data used for training, validating, and subsequent testing should be transparent for each AI algorithm to allow for independent testing with a previously unseen dataset avoiding potential bias by using repeated cases. The ground truth used should be reliable as it will greatly influence the algorithm.

Potential biases need to be considered before integrating AI-CAD into clinical practice. One of them is known as the "anchoring effect". Once markers for potential malignancy are placed on the image, readers tend to make their interpretations in relation to this initial reference and become biased in their decision making.[103] CAD systems are also known to lead to automation bias, where readers over-rely on the computer-generated decision by either committing false positives or omitting the undetected abnormalities.[104,105] Moreover, the "bandwagon effect" could occur with CAD systems based on supervised learning; readers may tend to align their opinion with the algorithm, knowing that the automated decision is the result of a large collection of annotations by experts.[106] In addition to training the AI-CAD system for improved predictions, there is also a need to train future radiologists to beware of such biases in the coming era of AI.

The European Union's General Data Protection Regulations[107,108] emphasise the importance of upholding stringent privacy and security regulations and AI-CAD systems should incorporate an element of explainable AI to enhance interpretability of an AI decision. Human-in-the-loop AI solutions are also an option that provides the opportunity

Please cite this article as: Le EPV et al., Artificial intelligence in breast imaging, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.02.006

for clinicians to interact with the model in order to improve the model's performance and to act as a safeguard from detrimental erroneous mistakes.

Further studies are needed to provide evidence for cost effectiveness due to the high cost of storage of data and use of GPUs. The acceptability of AI integration into workflows as well as how tools will deal with artefacts produced on clinical images requires further enquiry. Lastly the education required for radiologists before tools can be used in clinical practice should be assessed as well as potential recommendations for incorporating education on AI into radiologists training.

Future applications of algorithms evaluating risk prediction have the potential to inform surveillance and surgical decision making.[109] This will only be possible by the linkage of data from various sources including clinical information as well as genetic profiles and a collaboration between multiple healthcare fields with secure and interoperable data flows. The continuous ability for decision feedback and adjustment of an algorithm should be acknowledged in regulation and standardised performance audits.

AI-CAD tools: implementation in clinical practice

Software needs to be tested in clinical settings to prove the value in the clinical environment. Initial retrospective evaluation providing evidence of efficacy of algorithms is essential before moving into expensive clinical trials. A large and representative dataset is required, which includes all cancers from a large screening cohort when testing an algorithm to be used in screening mammography. Ideally, sequential cancers should be included together with the interval cancers and subsequent round cancers from that group of women, in addition to a large number of normal examinations representative of the screening population. Once the diagnostic accuracy is ascertained, comparisons can be made with different algorithms on a standardised independent test dataset. Algorithms proven to demonstrate efficacy can then be applied to prospective studies to test performance in day-to-day work flow. The reader's interaction with the tools is an important component as not all readers behave in a similar manner. A cost-effectiveness assessment should be undertaken to determine the impact on the health system in which it is being proposed, with a large feasibility study carried out to ensure there are no unintended consequences of introducing this new technology.

Opportunities for the application of AI in CAD mammography systems remain ripe, these could include: (1) a pre-screening triage system to allow radiologists to focus their time and energy on the more difficult and complex cases, or where there is uncertainty; (2) breast cancer risk prediction: incorporation of a multitude of clinical variables with imaging data to calculate *a priori* breast cancer risk; (3) automation of time-consuming subcomponents of mammography interpretation such as lesion segmentation. A successful AI-CAD tool is expected to be completely reliable in dismissing normal cases with highly confident scores, so that the radiologist can instead focus on cases that are more likely to be positive. The Holy Grail would be a tool to classify those cancers that are life-threatening requiring treatment and those that can be safely ignored to reduce over-diagnosis.

Conclusion

The transition from rule-based CAD systems to DL solutions with embedded domain knowledge has the potential to reduce diagnostic errors, improve radiologist accuracy, and help with decision-making.[53] DL advantages include the ability for end-to-end training and automatic hierarchical feature learning. Although there are several AI-CAD tools on the market today, innovation in ML and breast imaging will continue to drive their improvement. National and international collaborations between radiologists, computer scientists, academia, and industry are therefore necessary to regulate the integration of AI-CAD systems rapidly, safely, and effectively into clinical practice.

Conflict of interest

FJ Gilbert has a Consultancy arrangement with Alphabet Inc.

Acknowledgements

F.J.G. receives funding support for the National Institute for Health Research senior investigator award. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. E.P.V.L. is undertaking a PhD funded by the Cambridge School of Clinical Medicine, Frank Edward Elmore Fund and the Medical Research Council's Doctoral Training Partnership (award reference: 1966157). Y.W. and Y.H. are funded by the EPSRC Centre for Mathematical and Statistical Analysis of Multimodal Clinical Imaging (EP/N014588/1).

References

1. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pac J Cancer Prev* 2016;**17**:43–6.
2. DeSantis CE, Ma J, Goding Sauer A, *et al*. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* 2017;**67**:439–48.
3. Pharoah PDP, Sewell B, Fitzsimmons D, *et al*. Cost effectiveness of the NHS breast screening programme: life table model. *BMJ* 2013;**346**:f2618.
4. Lauby-Secretan B, Scoccianti C, Loomis D, *et al*. Breast-cancer screening — viewpoint of the IARC working group. *N Engl J Med* 2015;**372**:2353–8.
5. Marmot MG, Altman DG, Cameron DA, *et al*. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;**380**:1778–86.
6. Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;**339**:b2587. b2587.

7. Bond M, Pavey T, Welch K, *et al*. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess* 2013;**17**:1–170.
8. Tosteson ANA, Fryback DG, Hammond CS, *et al*. Consequences of false-positive screening mammograms. *JAMA Intern Med* 2014;**174**:954–61.
9. Elmore JG, Jackson SL, Abraham L, *et al*. Variability in interpretive performance at screening mammography and associated with accuracy. *Radiology* 2009;**253**:641–51.
10. Miglioretti DL, Smith-Bindman R, Abraham L, *et al*. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst* 2007;**99**:1854–63.
11. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *npj Breast Cancer* 2017;**3**:12.
12. Hoff SR, Abrahamsen AL, Samset JH, *et al*. Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography— results from a retrospective review. *Radiology* 2012;**264**:378–86.
13. Lehman CD, Wellman RD, Buist DS, *et al*. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015;**175**:1828.
14. Mitchell, T .M. *Machine Learning*. (McGraw-Hill, New York, USA, 1997). ISBN:0070428077 9780070428072
15. LeCun YA, Bengio Y, Hinton GE. Deep learning. *Nature* 2015;**521**:436–44.
16. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal* 2012;**16**:933–51.
17. Bengio Y, Lamblin P, Popovici D, *et al*. Greedy layer-wise training of deep networks. *Adv Neural Inf Process Syst* 2007;**19**:153.
18. Zhou Z-H. A brief introduction to weakly supervised learning. *Natl Sci Rev* 2018;**5**:44–53.
19. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;**20**:273–97.
20. Quinlan JR. Induction of decision trees. *Mach Learn* 1986;**1**:81–106.
21. Svetnik V, Liaw A, Tong C, *et al*. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003;**43**:1947–58.
22. Bengio Y. Deep learning of representations for unsupervised and transfer learning. *JMLR: Workshop Conf Proc* 2011;**7**:1–20.
23. Chartrand G, Cheng PM, Vorontsov E, *et al*. Deep learning: a primer for radiologists. *RadioGraphics* 2017;**37**:2113–31.
24. Dubrovina A, Kisilev P, Ginsburg B, *et al*. Computational mammography using deep neural networks. *Comp Meth Biomech Biomed Eng Imaging Vis* 2016;**6**:1–5.
25. Yamashita R, Nishio M, Do RKG, *et al*. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;**9**:611–29.
26. Mendel K, Li H, Sheth D, *et al*. Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography. *Acad Radiol* 2018, https://doi.org/10.1016/j.acra.2018.06.019.
27. U.S. Food and Drug Administration. *Summary of safety and effectiveness data: R2 technologies*. 1998P970058.
28. Nagi J, Abdul Kareem S, Nagi F, *et al*. Automated breast profile segmentation for ROI detection using digital mammograms. In: *Proceedings of 2010 IEEE EMBS conference on biomedical engineering and sciences, IECBES 2010 87–92*. IEEE; 2010, https://doi.org/10.1109/IECBES.2010.5742205.
29. Kegelmeyer WP, Pruneda JM, Bourland PD, *et al*. Computer-aided mammographic screening for spiculated lesions. *Radiology* 1994;**191**:331–7.
30. Zhang E, Wang F, Li Y, *et al*. Automatic detection of microcalcifications using mathematical morphology and a support vector machine. *Biomed Mater Eng* 2014;**24**:53–9.
31. Winsberg F, Elkin Jr M, Macy J, *et al*. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology* 1967;**89**:211–5.
32. El-Naqa I, Yongyi Yang Y, Wernick MN, *et al*. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging* 2002;**21**:1552–63.
33. Yassin NIR, Omran S, El Houby EMF, *et al*. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput Meth Progr Biomed* 2018;**156**:25–45.
34. Bowyer KW. *Digital image database with gold standard and performance metrics for mammographic. Image Analysis Research*. Tampa, USA: University of South Florida; 1998. Available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a366051.pdf. [Accessed 6 March 2019].
35. Lee RS, Gimenez F, Hoogi A, *et al*. Data Descriptor: a curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data* 2017;**4**:170–7.
36. Moreira IC, Amaral I, Domingues I, *et al*. INbreast: toward a full-field digital mammographic database. *Acad Radiol* 2012;**19**:236–48.
37. Gilbert FJ, Astley SM, Gillan MGC, *et al*. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;**359**:1675–84.
38. Taylor P, Potts HWW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 2008;**44**:798–807.
39. Becker AS, Marcon M, Ghafoor S, *et al*. Deep learning in mammography. diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Invest Radiol* 2017;**52**:434–40.
40. Kim E, Kim H, Han K, *et al*. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. *Nat Sci Rep* 2018;**8**:2762.
41. Chougrada H, Zouakia H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comp Meth Progr Biomed* 2018;**157**:19–30.
42. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal* 2017;**37**:114–28.
43. Nazari SS, Mukherjee P. An overview of mammographic density and its association with breast cancer. *Breast Cancer* 2018;**25**:259–67.
44. Morrish OWE, Tucker L, Black R, *et al*. Mammographic breast density: comparison of methods for quantitative evaluation. *Radiology* 2015;**275**:356–65.
45. Freer PE. Mammographic breast density: impact on breast cancer risk and implications for screening. *RadioGraphics* 2015;**35**:302–15.
46. Redondo A, Comas M, Macià F, *et al*. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol* 2012;**85**:1465–70.
47. Lehman CD, Yala A, Schuster T, *et al*. Mammographic breast density assessment using deep learning: clinical implementation. *Radiology* 2019;**290**:52–8.
48. Mohamed AA, Berg WA, Peng H, *et al*. Deep learning method for classifying mammographic breast density categories. *Med Phys* 2018;**45**:314–21.
49. Han Na L, Sohn Y-M. Mammographic density estimation by Volpara software: comparison with radiologists' visual assessment and relationship with BI-RADS category. In: *ECR Scientific Exhibit*; 2014, https://doi.org/10.1594/ecr2014/C-1957. C-1957.
50. U.S. Food and Drug Administration. *Quantra 510(k) K163623 clearance letter, December 20*. 2016. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163623.pdf. [Accessed 9 July 2018].
51. Rao AA, Feneis J, Lalonde C, *et al*. A pictorial review of changes in the BI-RADS Fifth Edition. *RadioGraphics* 2016;**36**:623–39.
52. U.S. Food and Drug Administration. *DM-Density 510(k) K170540 clearance letter, February 9, 2018*. 2018. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K170540.pdf. [Accessed 6 March 2019].
53. The digital mammography DREAM challenge. Available at: https://www.synapse.org/#!Synapse:syn4224222/wiki/401743. [Accessed 6 March 2019].
54. Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol* 2017;**3**:1463.
55. Gastounioti A, Conant EF, Kontos D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. *Breast Cancer Res* 2016;**18**:91.
56. Nikulin Y. Digital mammography DREAM challenge: participant experience 1. In *Medical imaging 2017: computer-aided diagnosis, 101344J*

10134. Bellingham, WA: SPIE. https://doi.org/10.1117/12.2280557. [Accessed 6 March 2019].
57. Russakovsky O, Deng J, Su H, *et al*. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015;**115**:211–52.
58. Carneiro G, Nascimento J, Bradley AP. *Unregistered multiview mammogram analysis with pre-trained deep learning models. Lecture notes in computer science*, vol. 9351. Cham: Springer; 2015. p. 652–60.
59. Basanth MS, Shettar R. Transfer learning on pre-trained deep convolutional neural network for classification of masses in mammograms. *IOSR J Comput Eng* 2017;**19**:50–5.
60. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012:1–9 https://doi.org/10.1016/j.protcy.2014.09.007.
61. DREAM Challenge results: can machine learning help improve accuracy in breast cancer screening? IBM Blog Research. Available at: https://www.ibm.com/blogs/research/2017/06/dream-challenge-results/. [Accessed 6 March 2019].
62. Patel MN Young K., Halling-Brown M. OPTIMAM Mammography Imaging Database (OMI-DB): a valuable dataset to fuel machine learning research In *SIIM 2017 scientific session analytics & deep learning Part 3*. Available at: https://cdn.ymaws.com/siim.org/resource/resmgr/siim2017/abstracts/analytics3-Patel.pdf. [Accessed 6 March 2019].
63. Llona Minguez, S. OPTIMAM mammography image database and viewing software. Available at: http://commercial.cancerresearchuk.org/sites/default/files/Optimam_May_2018.pdf. [Accessed 6 March 2019].
64. Kheiron Medical Technologies. Available at: https://www.kheironmed.com/. [Accessed 6 March 2019].
65. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;**25**:44–56.
66. Gilbert FJ, Tucker L, Gillan MG, *et al*. The TOMMY trial: a comparison of TOMosynthesis with digital mammography in the UK NHS breast screening programme — a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess* 2015;**19**:1–166.
67. Ciatto S, Houssami N, Bernardi D, *et al*. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;**14**:583–9.
68. Haas BM, Kalra V, Geisel J, *et al*. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 2013;**269**:694–700.
69. Reynolds A. Breast density and digital breast tomosynthesis. *Radiol Technol* 2013;**85**.
70. Tagliafico AS, Calabrese M, Bignotti B, *et al*. Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. *Eur Radiol* 2017;**27**:5179–84.
71. U.S. Food and Drug Administration. *Summary of safety and effectiveness data: PowerLook® Tomo detection software*. 2017P160009.
72. ScreenPoint Medical. Available at: https://www.screenpoint-medical.com/. [Accessed 6 March 2019].
73. Land C. ScreenPoint Medical showcases Transpara™, its machine learning software application designed to improve mammography reading efficiency and accuracy, at RSNA. ScreenPoint Medical press release. Available at: https://www.prweb.com/releases/2017/11/prweb14937787.htm. [Accessed 6 March 2019].
74. Hupse R, Samulski M, Lobbes MB, *et al*. Computer-aided detection of masses at mammography: interactive decision support versus prompts. *Radiology* 2013;**266**:123–9.
75. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, *et al*. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology* 2019;**290**:305–14.
76. The Royal College of Radiologists. *Guidance on screening and symptomatic breast imaging*. 3rd edition 2013.
77. Zanotel M, Bednarova I, Londero V, *et al*. Automated breast ultrasound: basic principles and emerging clinical applications. *Radiol Med* 2018;**123**:1–12.
78. QView Medical Inc. Announces FDA PMA approval for QVCAD™ for the GE Invenia™ 3D Automated Breast Ultrasound System ("ABUS"). Available at: https://www.businesswire.com/news/home/20171221006167/en. [Accessed 6 March 2019].
79. van Zelst JCM, Tan T, Clauser P, *et al*. Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. *Eur Radiol* 2018;**28**:1–11.
80. QView Medical. Available at: https://www.qviewmedical.com/. [Accessed 6 March 2019].
81. U.S. Food and Drug Administration. *Summary of safety and effectiveness data: QVCAD System*. 2016P150043.
82. Barinov L, Jairaj A, Paster L, *et al*. Decision quality support in diagnostic breast ultrasound through artificial Intelligence. In: *2016 IEEE signal processing in medicine and biology symposium, SPMB 2016*. Piscataway, NJ: IEEE; 2017, https://doi.org/10.1109/SPMB.2016.7846873.
83. Koios DS. AI-based ultrasound analysis. Available at: https://koiosmedical.com/solutions/. [Accessed 6 March 2019].
84. Podilchuk CI, Jairaj A, Barinov L, *et al*. Method and means of CAD system personalization to reduce intraoperator and interoperator variation. Appl No. 15/451,086 *US Patent Appl* 2016;**200**:266. Available at: https://patentimages.storage.googleapis.com/ee/06/69/f24bfe3a895128/US9536054.pdf [Accessed 6 March 2019].
85. Mann RM, Kuhl CK, Kinkel K, *et al*. Breast MRI: guidelines from the european society of breast imaging. *Eur Radiol* 2008;**18**:1307–18.
86. National Institute for Health and Care Excellence. *Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer*. NICE Guidance; 2017. Available at: https://www.nice.org.uk/guidance/cg164/ifp/chapter/early-detection-of-breast-cancer-by-surveillance. [Accessed 9 July 2018].
87. Sardanelli F, Fallenberg EM, Clauser P, *et al*. Mammography: an update of the EUSOBI recommendations on information for women. *Insights Imaging* 2017;**8**:11–8.
88. Monticciolo DL, Newell MS, Moy L, *et al*. Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. *J Am Coll Radiol* 2018;**15**:408–14.
89. Sardanelli F, Boetes C, Borisch B, *et al*. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. *Eur J Cancer* 2010;**46**:1296–316.
90. Pan J, Dogan BE, Carkaci S, *et al*. Comparing performance of the CADstream and the DynaCAD breast MRI CAD systems: CADstream vs. DynaCAD in breast MRI. *J Digit Imaging* 2013;**26**:971–6.
91. Meeuwis C, van de Ven SM, Stapper G, *et al*. Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0 T. *Eur Radiol* 2010;**20**:522–8.
92. Artificial intelligence to improve breast cancer screening. Available at: https://www.tees.ac.uk/sections/news/pressreleases_story.cfm?story_id=6784. [Accessed 6 March 2019].
93. First Option Software Limited. An enhanced artificial intelligence breast MRI scanning system (IntelliScan). Available at: http://gtr.ukri.org/projects?ref=104192. [Accessed 6 March 2019].
94. Aghaei F, Tan M, Hollingsworth A, *et al*. Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy. *Int Soc Magn Reson Med* 2016;**44**:1099–106.
95. Aghaei F, Tan M. Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy. *Med Phys* 2015;**42**:6520–8.
96. Braman NM, Etesami M, Prasanna P, *et al*. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res* 2017;**19**:1–14.
97. Ha R, Chang P, Karcich J, *et al*. Predicting post neoadjuvant axillary response using a novel convolutional neural network algorithm. *Ann Surg Oncol* 2018;**25**:3037–43.
98. Ha R, Chang P, Mutasa S, *et al*. Convolutional neural network using a breast MRI tumor dataset can predict oncotype Dx recurrence score. *J Magn Reson Imaging* 2019;**49**:518–24.
99. Ha R, Chang P, Karcich J, *et al*. Axillary lymph node evaluation utilizing convolutional neural networks using MRI dataset. *J Digit Imaging* 2018;**31**:851–6.
100. Li L, Roth R, Germaine P, *et al*. Contrast-enhanced spectral mammography (CESM) versus breast magnetic resonance imaging (MRI): a

retrospective comparison in 66 breast lesions. *Diagn Interv Imaging* 2017;**98**:113–23.
101. Fallenberg EM, Schmitzberger FF, Amer H, *et al*. Contrast-enhanced spectral mammography vs. mammography and MRI — clinical performance in a multi-reader evaluation. *Eur Radiol* 2017;**27**:2752–64.
102. Mendelson MB. Artificial intelligence in breast imaging: potentials and limitations. *AJR Am J Roentgenol* 2019;**212**:239–99.
103. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *RadioGraphics* 2015;**35**:1668–76.
104. Alberdi E, Povyakalo A, Strigini L, *et al*. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Acad Radiol* 2004;**11**:909–18.
105. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology* 2009;**253**:17–22.
106. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making* 2015;**35**:539–57.
107. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a 'right to explanation'. *AI Magazine* 2016;**38**, https://doi.org/10.1609/aimag.v38i3.2741.
108. EU GDPR Information Portal. Available at: https://www.eugdpr.org/. [Accessed 6 March 2019].
109. Bahl M, Barzilay R, Yedidia AB, *et al*. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology* 2018;**286**:810–8.

Please cite this article as: Le EPV et al., Artificial intelligence in breast imaging, Clinical Radiology, https://doi.org/10.1016/j.crad.2019.02.006