



## Short communication

## Artificial intelligence for breast cancer screening: Opportunity or hype?

Nehmat Houssami <sup>a,\*</sup>, Christoph I. Lee <sup>b</sup>, Diana S.M. Buist <sup>c</sup>, Dacheng Tao <sup>d</sup><sup>a</sup> Sydney School of Public Health, Sydney Medical School, University of Sydney, Australia<sup>b</sup> Department of Radiology, University of Washington School of Medicine, Seattle, WA, USA<sup>c</sup> Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA<sup>d</sup> UBTech Sydney AI Centre, School of Information Technologies, University of Sydney, Australia

## ARTICLE INFO

## Article history:

Received 29 August 2017

Accepted 12 September 2017

Available online 20 September 2017

## Keywords:

Artificial intelligence

Mammography

Population screening

## ABSTRACT

Interpretation of mammography for breast cancer (BC) screening can confer a mortality benefit through early BC detection, can miss a cancer that is present or fast growing, or can result in false-positives. Efforts to improve screening outcomes have mostly focused on intensifying imaging practices (double instead of single-reading, more frequent screens, or supplemental imaging) that may add substantial resource expenditures and harms associated with population screening. Less attention has been given to making mammography screening practice 'smarter' or more efficient. Artificial intelligence (AI) is capable of advanced learning using large complex datasets and has the potential to perform tasks such as image interpretation. With both highly-specific capabilities, and also possible un-intended (and poorly understood) consequences, this viewpoint considers the promise and current reality of AI in BC detection.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Breast cancer screening

Most developed health care systems have implemented population screening for breast cancer (BC) based on evidence from randomised trials that mammography confers BC mortality reduction, complemented by observational evidence of benefit from real-world screening [1,2]. BC screening involves interpretation of digital mammograms to identify suspicious abnormalities that warrant further investigation (recall to assessment) to rule in or rule out BC. Mammography interpretation is subjective. In addition to detecting BC, mammographic interpretation can yield false-positive results [3]; the frequency of false-positives varies between readers and screening programs and has been reported in the range of 3–12% of screening exams [4,5]; it is considerably higher when estimated as a cumulative false-positive rate allowing for repeat screening participation [3]. Human interpretation can also miss a BC – around 15%–35% of cancers occurring in screened women are not detected at imaging and present clinically as interval cancers [6] – due to either error or because the cancer is not visible or perceptible to the radiologist.

Much of the effort to improve population BC screening outcomes in recent decades has focused on intensifying screening practices, for example by using double-reading (2 independent interpretations) instead of single-reading, more frequent (annual vs biennial) screening, or incorporating new imaging technologies to enhance cancer visualisation (tomosynthesis) [7] with substantial resource requirements, or at the cost of yielding more false-positives (supplemental ultrasound and MRI) [8,9]. Little attention has been given to making screening practice using mammography simply 'smarter' and potentially more efficient. Within this complex landscape of BC screening, artificial intelligence (AI) may mistakenly seem akin to computer-aided detection, but unlike the latter and its excessive false-prompting of the human reader, AI is capable of advanced learning and has the potential to one day perform tasks such as stand-alone interpretation [10]. With highly-specific capabilities, and also potential un-intended consequences [11], both the promise and reality of AI in BC detection must be considered.

## 2. Artificial intelligence (AI)

Machine learning (ML), a rapidly growing field of AI, integrating computer science and statistics, allows computers to learn without

\* Corresponding author. Sydney School of Public Health (A27), Sydney Medical School, University of Sydney, Sydney 2006, Australia.

E-mail address: [nehmat.houssami@sydney.edu.au](mailto:nehmat.houssami@sydney.edu.au) (N. Houssami).

explicit programming through automatic extraction and analysis of complex data [12–17]. ML is being touted as a potential AI tool to help discover new materials, master various games, and improve predictive ability in clinical medicine. The convergence of new ML techniques, such as deep learning (a class of learning algorithms that stacks a set of non-linear operations to extract features and explore transformations from training data) and reinforcement learning (similar to training by rewards, this allows ML to tackle a problem involving sequential decisions, to determine the ideal actions within a specific context that maximize rewards and performance), current computer processing capabilities, and the rapid growth of digital capture and storage of data in general and specifically in science and health, are transforming the application of AI in diverse areas. For example, Raccuglia et al. [16] used the processed unsuccessful hydrothermal synthesis as negative examples to train support vector machines (a supervised learning model). The model was then utilized to predict the reaction of success/failure of a new input synthesis, and successful synthesis further verified by human capability. The rate of a successful synthesis recommended by the model (and verified to be a true success) was significantly higher than that of a synthesis recommended by a human's intuition. This AI strategy could be applied to enhance discovery of structures associated with cancers that were missed in routine screening practice. A new deep learning architecture was recently proposed to significantly reduce the number of data required to train a predictive model for drug discovery [12], further enhancing the feasibility and potential role of AI in health research.

The AI field is growing rapidly in various applications through advanced computational and statistical sciences. Discriminative Gaussian Process models [15] trained on 'human face' data from multiple sources have now surpassed human-level performance on image-based face recognition. Trunk-branch ensemble convolutional neural network (TBE-CNN) trained on both still face images and simulated video data has been shown to outperform the averaged human level performance in recognizing dynamic face images captured under surveillance circumstances [13]. CNNs trained on large amounts of labelled clinical images have been utilized to automatically classify skin lesions [14]: using lesion classification results, a 'skin cancer' category could be identified at a performance level that is comparable to that of dermatologists. In other applications, deep learning has also been combined with reinforcement learning to master the challenging game of Go [17]; the developed program AlphaGo achieved a complete victory in this game versus the human European Go champion.

### 3. Exploration of AI for BC screening

Medical imaging is ripe for incorporating AI solutions because of existing capabilities in computer vision and automated image feature analysis, the large storage capacity for hundreds of thousands of digital imaging exams through picture archiving and communication systems (PACS), linkage of PACS to electronic medical records, and the binary outcome of imaging-based screening tests [10]. Recent work developing AI algorithms in digital mammography interpretation has undertaken the conversion of single whole digital images of the breast into automatically extracted quantitative, pixel-level variables [10]. Using ascertained mammographic examinations, computers can cluster these millions of pixel-level variables unrecognizable by the human eye to identify new imaging features associated with BC, which can be associated with gold-standard outcome data for training. In addition, AI can further combine pixel-level variables and associations with patient-level clinical data, including known patient risk factors, to develop sophisticated predictive algorithms that may one day equal or outperform human screening mammography

interpretive accuracy [10].

The ongoing *Digital Mammography DREAM Challenge* is one of the larger efforts in using AI to attempt improving BC screening outcomes [10]. This crowdsourcing coding competition, offering a large monetary prize for the best algorithm for predicting BC on screening mammography, has brought together over 1200 coding teams worldwide to develop an AI algorithm that can detect BC with accuracy almost equivalent to that of radiologists, with the goal of enhancing accuracy through decreasing the recall rate from BC screening [10]. The ongoing collaborative phase where the best teams (the challenge 'winners') work together to improve the final algorithm holds promise for improved accuracy that may detect BC with accuracy that may rival that of radiologists. Final Challenge results are expected late 2017 with open access to the winning coding algorithms. Hence, ongoing short to intermediate term activities will focus on improving this open-source algorithm to achieve higher accuracy, building on the Challenge results to bring AI products to market, and exploring how an accurate algorithm might be potentially incorporated into screening practice.

### 4. Future research using AI for BC screening

Although research into the capability of AI in mammography screening at present focuses on interpretive accuracy with the anticipation of primarily reducing false-positives, there remains much opportunity to extend exploration of the role of AI in BC screening, for example to develop future models that also better identify BC. This would not be limited to potential detection of cancers 'missed' at human screen-reading (with the aim of reducing the frequency of interval cancers) but may potentially improve detection of more aggressive forms of BC or faster growing cancers with the possibility of enhancing screening benefit. AI models may also pave the way to differentiate aggressive from indolent screen-detected BC. Therein lies the possibility of mitigating the risk of over-diagnosis [5] from population BC screening through application of AI, although research into that very challenging issue will most likely require sourcing more information than what can be learned from imaging, such as developing models using image-level plus genomic-level variables.

### 5. Unexplored aspects of AI for BC screening

A number of unexplored issues, highly relevant for the application of AI in BC screening, warrant consideration and research effort. These include the social and ethical concerns and implications inherent in entrusting cancer detection to an AI model, and the possibility of un-intended consequences [11]. These issues need to be explored early in the phase of developing AI models for BC screening to provide an understanding of societal perspectives, and to define an ethical-legal framework for the potential application of AI models in cancer screening practice; this would also help ensure that the purpose for which AI models are developed is acceptable to all stakeholders. Un-intended consequences of using AI for BC screening may relate to screening outcomes (such as the possibility of increased detection of indolent carcinoma in-situ), or may represent consequences to breast imagers requiring re-skilling of readers and modifying the role of breast imagers in the BC screening and diagnostic continuum.

### 6. Conclusion

Whether the anticipated promise of AI in BC detection, or indeed more broadly in cancer research, translates into practice and meets expectations remains to be seen over the coming years. At present, however, AI represents an opportunity that is both feasible

and timely for exploration in population BC screening. To make the most of this opportunity, extremely large data-sets of imaging examinations linked to clinical factors and cancer outcomes are needed to train and validate robust AI models that can eventually be incorporated into clinical practice. Key to making progress will be a highly collaborative inter-disciplinary, inter-institutional, and international approach.

#### Financial disclosure

None to disclose.

#### Conflict of interest statement (COI)

None to declare.

#### Acknowledgements

N. Houssami receives research support through a National Breast Cancer Foundation (NBCF Australia), Breast Cancer Research Leadership Fellowship.

#### References

- [1] Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening — viewpoint of the IARC working group. *N Engl J Med* 2015;372(24):2353–8.
- [2] Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *The Lancet* 2012;380(9855):1778–86.
- [3] Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive services task force recommendation. *Ann Intern Med* 2016;164(4):256–67.
- [4] Breast Cancer Surveillance Consortium (BCSC). Performance measures for 1,838,372 screening mammography examinations from 2004 to 2008 by age (based on BCSC data through 2009). [Accessed 12 August 2017].
- [5] Houssami N. Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? *Cancer Biol Med* 2017;14(1):1–8.
- [6] Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *Nat (NPJ) Breast Cancer* 2017;3(1):12. . [Accessed 13 April 2017].
- [7] Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;14(7):583–9.
- [8] Tagliafico AS, Calabrese M, Mariscotti G, Durando M, Tosto S, Monetti F, et al. Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: interim report of a prospective comparative trial. *J Clin Oncol* 2016;34(16):1882–8.
- [9] Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. Supplemental screening for breast cancer in women with dense breasts: a systematic review for the U.S. Preventive services task force. *Ann Intern Med* 2016;164(4):268–78.
- [10] Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol* May 2017 [Epub ahead of print].
- [11] Cabitza F, Rasoini R, Gensini G. Unintended consequences of machine learning in medicine. *JAMA* 2017;318(6):517–8.
- [12] Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. *ACS Central Sci* 2017;3(4):283–93.
- [13] Ding C, Tao D. Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Trans Pattern Anal Mach Intell* 2017;PP(99):1.
- [14] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115–8.
- [15] Lu C, Tang X. Surpassing human-level face verification performance on LFW with GaussianFace. In: Proceedings of the 29th AAAI conference on artificial intelligence (AAAI-15); 2014.
- [16] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* 2016;533(7601):73–6.
- [17] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;529(7587):484–9.