European Journal of Radiology 167 (2023) 111087

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Overview of trials on artificial intelligence algorithms in breast cancer screening – A roadmap for international evaluation and implementation

T.J.A. van Nijnatten [a,b,c], N.R. Payne [a], S.E. Hickman [a,d], H. Ashrafian [e], F.J. Gilbert [a,f,*]

[a] Department of Radiology, University of Cambridge School of Clinical Medicine, Box 218, Level 5, Cambridge Biomedical Campus, Cambridge CB2 0QQ, United Kingdom
[b] Department of Radiology and Nuclear Medicine, Maastricht University Medical Center+, Maastricht, the Netherlands
[c] GROW – School for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, the Netherlands
[d] Department of Radiology, Barts Health NHS Trust, The Royal London Hospital, 80 Newark Street, London E1 2ES, United Kingdom
[e] Institute of Global Health Innovation, Department of Surgery and Cancer, Imperial College London, St Mary's Hospital, London, United Kingdom
[f] Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, United Kingdom

ARTICLE INFO

Keywords:
Breast imaging
Screening
Artificial Intelligence
Algorithms
Evaluation

ABSTRACT

Accumulating evidence from retrospective studies demonstrate at least non-inferior performance when using AI algorithms with different strategies versus double-reading in mammography screening. In addition, AI algorithms for mammography screening can reduce work load by moving to single human reading. Prospective trials are essential to avoid unintended adverse consequences before incorporation of AI algorithms into UK's National Health Service (NHS) Breast Screening Programme (BSP). A stakeholders' meeting was organized in Newnham College, Cambridge, UK to undertake a review of the current evidence to enable consensus discussion on next steps required before implementation into a screening programme.

It was concluded that a multicentre multivendor testing platform study with opt-out consent is preferred. AI thresholds from different vendors should be determined while maintaining non-inferior screening performance results, particularly ensuring recall rates are not increased. Automatic recall of cases using an agreed high sensitivity AI score versus automatic rule out with a low AI score set at a high sensitivity could be used. A human reader should still be involved in decision making with AI-only recalls requiring human arbitration. Standalone AI algorithms used without prompting maintain unbiased screening reading performance, but reading with prompts should be tested prospectively and ideally provided for arbitration.

1. Introduction

National breast screening programmes have been successfully implemented for early breast cancer detection to facilitate optimal treatment possibilities with more favourable prognostic outcome [1–3]. Programmes vary in their structure with double or single reading of cases, the use of digital mammography or breast tomosynthesis as well as the age of invitation. Recently, strategies have been proposed to implement artificial intelligence (AI) algorithms into mammography screening programmes with AI replacing one of the two readers, with AI triaging cases to be single or double read, and AI identifying cases to be recalled [4]. Multiple independent retrospective studies have demonstrated that the use of AI algorithms to replace one reader in screening mammography has acceptable performance compared to double-reading [5–9]. In addition, AI algorithms in screening mammography have the potential to significantly reduce workload and decrease the number of false positive recalls [10–13].

A recent report from the UK National Screening Committee concluded that incorporation of any AI algorithm into the breast cancer screening programme requires prospective studies to obtain key evidence to measure the effect of AI algorithms on human readers and the impact on patients' outcomes [14].

A stakeholders' meeting was organized in Newnham College, Cambridge, UK to review the current retrospective evidence and ongoing and planned prospective international trials to enable consensus discussion on next steps required before implementation into the UK's National Health Service Breast Screening Programme (NHSBSP). Discussion was held about the type of prospective trial that could be undertaken in the

* Corresponding author at: Department of Radiology, University of Cambridge School of Clinical Medicine, Box 218, Addenbrookes Hospital, Cambridge Biomedical Campus, Cambridge CB2 0QQ, United Kingdom.
E-mail address: fjg28@cam.ac.uk (F.J. Gilbert).

https://doi.org/10.1016/j.ejrad.2023.111087
Received 12 July 2023; Received in revised form 23 August 2023; Accepted 4 September 2023
Available online 8 September 2023

UK to complement the initial prospective work being undertaken.

Presentations were given by speakers giving the UK and international perspectives, the UK breast screening programme and companies with AI tools ready for testing or already being implemented in international screening programmes. Consensus statements were defined and voted by all those in attendance at the meeting's conclusion.

2. Stakeholders' meeting

This meeting was held on October 4th 2022, XX and was organized by XX and XX. Invitations were sent to AI companies with a viable tool who had expressed an interest in working in the UK, principal investigators of published or ongoing mammography AI trials, members of the UK national screening committee and representatives of the breast screening programme and policy makers. Delegates from Denmark, the Netherlands, Norway, South Korea, Sweden and USA attended of whom 61 attendees, 53 were on-site and 8 were online (Table 1).

3. Review of retrospective evidence

Retrospective studies have explored various applications of AI in breast cancer screening programmes across the globe, with an acceleration in evidence since the 2016 Digital Mammography (DM DREAM) challenge [5–7]. The retrospective studies have evaluated AI in as a stand-alone reader to provide an independent screen reader decision, going one step further than traditional computer aided detection (CAD) methods where a prompt is provided on an image to draw the reader's attention to an abnormality [5,6]. A systematic review of five small studies (for use of AI as stand-alone reader or as a reader aid) demonstrated that the AI algorithms "were more accurate than a single radiologist reading a test set in the laboratory" [6]. However, when analysing three larger studies 94.0 % of AI algorithms were "less accurate" when compared to the single "original radiologist" performance [6]. A further systematic review found an AUC of 0.89 and 0.85 for AI and readers respectively, although there was no statistically significant difference in performance of the pooled sensitivity and specificity results for AI algorithms compared to readers when investigating studies under set conditions (e.g. top-performing algorithm tested on external data set) [5]. Simulation studies have reported non-inferiority of AI used in combination with the original single reader decision when compared to double reading, and some studies have reported superior specificity with this approach [10,15]. However, there are numerous sources of bias in retrospective study design such as small sample size, enriched test sets, lack of external datasets, as well as differential verification bias of the reference standard [5,6,14]. Furthermore, it is not known how a reader would respond to an AI decision in "real life", what arbitration decisions would be made with AI discrepancies, including if AI location prompts were reviewed by readers. This is the shortcoming of simulation studies and should be taken into account when evaluating results of retrospective studies [6,14,16].

An alternative application demonstrating promise is the use of AI as a triage tool [5,6]. AI triage can be used at one end of the spectrum by safety-netting high suspicion cases for either prioritized review, enhanced assessment, or automatic recall [12,17]. Alternatively at the other end of the spectrum AI triage can be used to provide a final decision for very low suspicion cases which could mean they are either read by one reader or do not require a further read [10,12,17–21]. Results from retrospective studies have shown between 17.0 % and 91.0 % of cases can be triaged as a low suspicion to be read by an adapted method such as a single reader [5,6]. A further advantage of this approach is the reduction of false positive cases [20]. The number of cancers not detected by the AI when used to identify "normal" ranges between 0.0 % and 10.3 % [10,12,17–21]. Studies investigating triage of high suspicion cases have demonstrated that between 12.0 % and 48.0 % interval cancers could be detected [12,22–26]. Additionally, between 14.0 % and 45.0 % next round cancers could be detected at the earlier screen [12]. However, only a small proportion of cases can be classified and recalled for further enhanced assessment due to both the high cost and time intensity of assessment clinics, limiting the number of interval and next round cancers that would actually be detected [12]. Retrospective studies have also examined both ends of the AI triaging spectrum together ("rule in" and "rule out"), finding a significant reduction in workload whilst achieving either non-inferior or improvements in sensitivity [11,17,27].

Over recent years the amount of retrospective evidence has increased, as well as there have been improvements in the size and quality of data used for evaluation. Overall retrospective studies have a place in the evaluation pathway of AI, enabling scientists to test various applications, using different commercial and academic algorithms, on different cohorts of patients as well as identifying the appropriate threshold for use. Thus, retrospective studies provide a framework for prospective work and guided programme adoption.

4. Ongoing prospective trials

Ongoing prospective trials from around the globe on the topic of AI algorithms in screening mammography were presented and discussed. An overview of the completed prospective trials is provided in below, including proposed AI strategy and the used AI thresholds used in each study (Table 2).

4.1. Denmark

Initially, a retrospective simulation study from the Danish Capital Region breast cancer screening program demonstrated that in a cohort consisting of 114,421 women (age 50–69 years) an AI algorithm had non-inferior sensitivity (69.7 % versus 70.8 %, $p = 0.02$) and higher specificity (98.6 % versus 98.1 %, $p = 0.001$) compared to the double reading results from the radiologists [11]. In addition, radiologist workload was calculated to be reduced by 63 % because of the AI algorithm and 25 % of the false-positive screenings would have been avoided.

The Danish Capital Region breast cancer screening programme decided to adopt this AI algorithm in clinical practice and evaluate it on an ongoing basis (MAGIC – MammGraphy AI in Breast cancer diagnostics). An AI score 1–5 is single-read, AI score 6–10 requires double-reading, and where the AI score is > 9.98 the woman will be recalled to assessment based on AI only. Preliminary results presented at the meeting are very encouraging.

4.2. Germany

The PRAIM Study (PRospective multicentre observational study of an integrated Artificial Intelligence (AI) system with live Monitoring) is ongoing to prospectively investigate double-reading versus single-reading + AI algorithm, within a cohort of 400,000 women (age 50–69 years) participating in the German Mammography Screening

Table 1
Primary profession of attendees.

Profession	Number of attendees	Geographical region
Breast radiologist	16	Europe, South East Asia, USA
Epidemiologist	1	UK
UK NHSBSP staff/Health policy staff	9	UK
AI Commercial Vendor	18	Europe, South East Asia, USA
Research personnel	6	UK
Miscellaneous, including patients	11	Europe, South East Asia, USA

Table 2
Overview of ongoing prospective trials on AI algorithms in mammography screening.

Trial name (ID)	Study design	Country	Principal Investigator	Number of participants	AI vendor	AI strategy
n/a*	Clinical evaluation	Denmark	I. Vejborg	n/a*	Transpara (Screenpoint Medical)	AI score determines number of readers: score 1–5 = SR, score 6–10 = DR score > 9.98 = recall by AI
PRAIM (DRKS00027322)	Observational	Germany	A. Katalinic	400,000	Vara	DR versus SR + AI
AI-STREAM (NCT05024591)	Observational	South Korea	Y. Chang	32,714	Lunit	SR versus SR + AI
AITIC (NCT04949776)	Observational	Spain	E. Cabot	27,000	Transpara	AI score determines number of readers: Score < 8: none Score > 7: DR + AI assist
MASAI (NCT04838756)	RCT	Sweden	K. Lång	100,000	Transpara	DR versus SR + AI
ScreenTrusCAD (NCT04778670)	Observational	Sweden	F. Strand	55,579	Lunit	DR versus SR + AI versus independent AI
AI-ROL (NCT05048095)	Observational	Sweden	H. Gustafsson	15,500	Transpara	DR + AI Cases not recalled by DR but 3 % most suspicious AI score: recall by AI
GEMINI (NHSX phase 3)	Prospective	United Kingdom	G. Lip	200,000	Kheiron	DR with AI in background
Service evaluation (NHSC phase 4)	Observational	United Kingdom	Various PI's	Unknown	Kheiron	AI in silent background
AIMS	Prospective	United Kingdom	A. Darzi	Unknown	Google	AI in silent background

Abbreviations: *ID* identification number, *AI* artificial intelligence, *n/a* = not applicable, *RCT* randomized controlled trial, *DR* double reading, *SR* single reading, *Transpara score 1–5* = 50 % of cases with relatively low risk breast cancer score, *Transpara score 6–10* = 50 % of cases with relatively high risk breast cancer score.
* AI already incorporated in nationwide breast cancer screening.

Programme [28]. The decision-referral combines triage of cases to normal examinations and those with high likelihood of cancer with a collaborative human-AI approach. Non-inferiority is evaluated with a weighted, mixed-effects linear regression model and defined as −10 % at the lower bound of the two-sided 95 % confidence interval which corresponds to +/− 0.6/1000 screen detected cancers. Recognising the variation between cancer centres the previous five years of cancer detection rate (CDR) is used. The live monitoring uses current and historical "controls", allows for greater participation of a diverse population across the German screening programme, and should produce rapid results of this evolving technology.

4.3. South Korea

The AI-STREAM (Artificial Intelligence for breast cancer screening in mAMmography) study is a prospective trial investigating single-reading versus single-reading + AI algorithm (computer aided detection/diagnosis) in a cohort of 32,714 women (age 40–80 years) participating in Korean breast cancer screening programme at five sites [16]. This trial will report the diagnostic accuracy of radiologists reading with and without CAD prompts. The reader first reads the mammogram unaided and records their decision and the re-reads the case with the prompts visible and decides whether or not to recall. Standalone AI will be tested in the background. Arbitration will also be examined separately where a third reader will decide without AI prompts.

4.4. Spain

A recent retrospective study consisting of mammography exams from the Córdoba Tomosynthesis Screening Trial demonstrated that an AI algorithm reduced workload up to 70 % without reducing sensitivity by 5 % or more [29]. From the results of this study, the AITIC trial (Artificial Intelligence in Breast Cancer Screening Programs in Córdoba) was designed. This prospective trial among 27,000 women (age 50–69 years) will investigate double-reading versus reading strategy based on provided score from the AI algorithm: score < 8 (low probability of cancer) will not be evaluated by any radiologist, score > 7 double-reading. The primary outcome measure is workload calculated multiplying the average time for a reading by the total number of readings, cancer detection rate and referral for further work-up. The secondary outcomes are PPV for referrals, biopsies and Transpara scores.

4.5. Sweden

In Sweden there are currently three independent ongoing prospective trials investigating incorporation of AI algorithms in mammography screening. The MASAI (Mammography Screening With Artificial Intelligence) study is a randomized-controlled trial investigating among 100,000 women (age 40–74 years) double-reading (control arm) versus single-reading + AI algorithm (intervention arm). In the case of AI score 1–9 single-reading will be performed, AI score 10 will require double-reading. Results of the study demonstrated a higher recall rate in the intervention arm of 2.2 % versus 2.0 %, including a higher cancer detection rate in the intervention arm (6.1 ‰ versus 5.1 ‰, P = 0.052) [30].

The second trial is the ScreenTrustCAD (Artificial Intelligence in Large-scale Breast Cancer Screening) observational study investigating double-reading versus single-reading + AI algorithm versus AI alone (only for secondary endpoints: reader flagging, consensus recall, process failure) among 55,579 women (age 40–74 years). Preliminary results presented at the annual meeting of Radiological Society of North-America in 2022 showed that combining the evaluations of one radiologist with AI compared to two radiologists improved cancer detection rate (relative true positive fraction 1.06 (95 % confidence intervals: 1.01–1.10) without an increase in recall rate (28.0/1000 versus 29.3/1000) [31]. Therefore, it was concluded that replacing one radiologist with AI is ready for clinical implementation in mammography screening programs.

The third trial is the observational AI-ROL (Artificial Intelligence in Breast Cancer Screening in Region Östergötland Linkoping) study investigating the use of AI algorithm as third reader and a decision support system during consensus in a cohort of 15,500 women (age 40–74 years). The primary outcome measures are CDR, referral rate and PPV of referrals with PPV of Transpara scores as a secondary measure.

4.6. United Kingdom

In the United Kingdom the Life Sciences Innovate UK funded two AI companies – Kheiron and Alphabet Inc to undertake prospective testing in NHS screening. The Kheiron retrospective trials 1 and 2 in Hungary and England confirmed the utility of AI and underpinned the CE mark with a further study undertaken in Scotland [32,33]. The ARIES multi-site retrospective study is being analysed. The prospective LIBRA is a single site paired study with AI as a third reader and used in arbitration with no prompts. The GEMINI service evaluation is comparing standard double-reading and AI as a third reader and consensus using prompts in Aberdeen, Scotland with a larger multi-site evaluation in England with 200,000 women.

Google is working on two studies focused on different use cases. In the US, a study has been completed at Northwestern in Chicago looking at the feasibility of using AI to triage women with suspicious lesions for same-day diagnostic workup. In the UK, the AI in Mammography Screening (AIMS) study is currently running at two NHS Trusts to determine the feasibility and impact of introducing AI as a second reader within a double-reader workflow, and to understand the important aspects of human–computer interaction when breast cancer specialists in an arbitration panel include AI outputs in their decision making. Results from this study are expected in 2024.

5. Consensus agreement at the meeting

A session was held proposing elements for a future prospective study to obtain key evidence required for implementation of AI algorithms into NHSBSP (Table 2). Various aspects were discussed and finally statements were given to the audience in order to establish whether or not consensus could be obtained among the participants (Table 3). Consensus was considered when >60 % of the on-site participants agreed on the proposed statement.

6. Proposed prospective study for the UK

To collect key evidence for incorporation of an AI algorithm in NHSBSP, elements for a proposed prospective study design were suggested at the stakeholders' meeting:

6.1. Algorithm factors

- Task identified
  - Triage (to rule in highly suspicious cases and to rule out definite normal cases)
  - Stand-alone AI alongside single human reader
  - Arbitration/consensus assessment
  - AI reading strategy: AI triaging, single-reading + AI algorithm or double reading with AI prompts. Consider automatic recall in the case of a high AI score versus automatic rule out in the case of a low AI score set at a high sensitivity threshold. Multiple AI reading strategies should be considered.
- Evaluated retrospectively
  - For an AI tool to be used in a prospective trial, benchmarking data needs to be available, tested on an independent UK screening dataset
  - Retrospective assessment of the AI should be non-inferior to single reader UK performance
  - Operating point for task and algorithm identified for prospective trial

    AI threshold: different cut-off values for AI score should be determined in order to maintain sensitivity, specificity, cancer detection rate, interval cancer rate and recall rate.
- Number of algorithms
  - Multi-vendor:

    All AI companies meeting UK standards should be able to participate

Table 3
Overview of consensus agreement and disagreement according to the meeting.

Consensus agreement	Consensus disagreement
We support a prospective service evaluation approach to investigate AI algorithm incorporation in mammography screening	We are ready for incorporation of autonomous AI algorithms in mammography screening
AI algorithms in mammography screening should be benchmarked for inclusion	We need to perform a randomized-controlled trial for AI algorithm trials in mammography screening

6.2. Study factors

- Design
  - It was recognized that this is dependent on the above task being evaluated and the time to output, discussion focused around the fast-moving field and what outcomes can be obtained as part of an initial trial as opposed to a lengthy RCT trial, however the level of evidence to ensure patient outcomes is also important
  - Randomized-controlled trial would be ideal, but this is a lengthy process and the danger is that the results would be outdated by time of reporting. RCT could delay adoption of AI into breast screening.
  - Observational study/evaluation would be the fastest method to start using AI but would require careful regular audits to ensure no adverse impact on patient care. This would start with background evaluation of AI with ongoing double reading. The next step would be AI in assessment and acting on any cases where AI was above a given threshold, while double reading was negative. Then move to single reading with AI which can be compared with historical data.
- Multicentre
  - Preferably at least 10 sites multivendor testing platform.
  - Statistical power with sufficient sample size in order to detect key outcomes
- Opt-out consent
  - Study information will be sent to women with no informed consent required but with the opportunity to opt-out.
- Mammography reading: to maintain unbiased mammography reading, AI algorithms should be used without prompting during mammography reading, except at consensus meeting. Recalls by AI might require arbitration to keep assessments as low as possible, auto-recall by AI might be considered given a high threshold. An AI arm readers using prompts should be considered.
- Outcome measures:
  - Sensitivity, specificity, cancer detection rate (including tumour size and cancer types), recall rate, arbitration rate
  - Workload using reading time as a measure

6.3. Infrastructure factors

- NHSBSP has already created a platform to allow testing of AI algorithms
- Adaptation to existing IT systems (NBSS) for data collection
- Incorporation of AI algorithm and possible storage of AI outputs into PACS
- Training readers on using AI tool. This will need to be done for each company's algorithm.

7. Conclusion

Breast cancer screening programmes are likely to change in the next few years regarding the mammography reading strategy, because of previous retrospective studies demonstrating at least non-inferior

performance when using an AI algorithm [5]. Several strategies have been proposed to change mammography screening, including AI serving as one of the two readers, AI selecting cases to be single or double read, and AI selecting cases to be recalled [4].

The UK National Screening Committee highlighted the key areas where evidence is required to measure the effect of implementation of AI algorithms within the NHSBSP. The stakeholders concluded that a multicentre multivendor testing platform with opt-out consent is required and a service evaluation was preferred to a full randomised controlled trial. The stakeholders agreed that AI algorithms should be benchmarked before inclusion and that autonomous AI was not appropriate at present. AI operating thresholds should be determined while maintaining non-inferior screening performance results and keeping recall within acceptable limites. Automatic recall in the case of a high AI score can be considered with automatic rule out in the case of a low AI score set at a high sensitivity. All AI-only recalls require arbitration. AI algorithms should be used without prompting initially at mammography reading to maintain unbiased screening reading performance, with prompts used for arbitration. An AI reader assisted arm with prompts could be included.

CRediT authorship contribution statement

T.J.A. van Nijnatten: . **N.R. Payne:** . **S.E. Hickman:** . **H. Ashrafian:** Writing – review & editing, Supervision, Investigation, Conceptualization. **F.J. Gilbert:** .

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Support for this meeting was from Newnham College, University of Cambridge, Cambridge, UK. F.J.G. is supported by a Cancer Research UK early detection programme grant. H.A. is Chief Scientific Officer, Preemptive Medicine and Health, Flagship Pioneering.].

Acknowledgements

The authors wish to thank all attendees for the interesting presentations and discussions at the stakeholders' meeting on October 4th 2022, Newnham College, Cambridge, United Kingdom. They are:

Speakers: F.J. Gilbert, H. Ashfarian, S. Hickman, F. Strand, K. Dunbar, S. Wright, E. Pisano, G. Lip, I. Vejborg, M. Halling-Brown, A. Brentnall, M. Richards, L. Wilkinson, S. Duffy.

Companies: D. Tse, R. Chopra, C. Kelly (Google), P, Kecskemethy, J. Nash (Kheiron Medical), C. Leibig, D. Byng (VARA), N. Karssemeijer, N. Janssen (Screenpoint Medical), S. Pacile, P. Fillard (Therapixel), K. Kim, J. Ahn (Lunit), S. Lee, R. Caponigro, J. Go, M. Debain, J. Hoffmeister (ICad), C. Austin, O. Charles (AWS)

Radiologists: N. Sharma, S. Vinnicombe, K. Satchithananda, L. Wilkinson, R. Given-Wilson, S. Muthyala, R. Sidebottom, A. Juette, M. Nanaa, M. Wallis, E. Pisano.

PPI: D. Keatly.

Funders: D. Crosby.

Various experts: Y. Chen, D. Cushnan, S. Webb, S. Duffy, B. Kasmai, G. Montana, E. Atakapa, J. Dunn, O. Kearins, S. McIntosh, J. Hall, Y. Chang.

Cambridge: D. Kessler, M. Graves, N. Payne, T. van Nijnatten, Y. Huang

References

[1] F. Sardanelli, H.S. Aase, M. Álvarez, E. Azavedo, H.J. Baarslag, C. Balleyguier, P. A. Baltzer, V. Beslagic, U. Bick, D. Bogdanovic-Stojanovic, R. Briediene, B. Brkljacic, J. Camps Herrero, C. Colin, E. Cornford, J. Danes, G. de Geer, G. Esen, A. Evans, M.H. Fuchsjaeger, F.J. Gilbert, O. Graf, G. Hargaden, T.H. Helbich, S. H. Heywang-Köbrunner, V. Ivanov, Á. Jónsson, C.K. Kuhl, E.C. Lisencu, E. Luczynska, R.M. Mann, J.C. Marques, L. Martincich, M. Mortier, M. Müller-Schimpfle, K. Ormandi, P. Panizza, F. Pediconi, R.M. Pijnappel, K. Pinker, T. Rissanen, N. Rotaru, G. Saguatti, T. Sella, J. Slobodníková, M. Talk, P. Taourel, R.M. Trimboli, I. Vejborg, A. Vourtsis, G. Forrai, Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey, Eur. Radiol. 27 (7) (2017) 2737–2743.

[2] S. Saadatmand, R. Bretveld, S. Siesling, M.M. Tilanus-Linthorst, Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients, BMJ 351 (2015), h4901.

[3] D.A. Berry, K.A. Cronin, S.K. Plevritis, D.G. Fryback, L. Clarke, M. Zelen, J. S. Mandelblatt, A.Y. Yakovlev, J.D. Habbema, E.J. Feuer, Effect of screening and adjuvant therapy on mortality from breast cancer, N. Engl. J. Med. 353 (17) (2005) 1784–1792. I. Cancer, C. Surveillance Modeling Network.

[4] M. Larsen, C.F. Aglen, S.R. Hoff, H. Lund-Hanssen, S. Hofvind, Possible strategies for use of artificial intelligence in screen-reading of mammograms, based on retrospective data from 122,969 screening examinations, Eur. Radiol. 32 (12) (2022) 8238–8246.

[5] S.E. Hickman, R. Woitek, E.P.V. Le, Y.R. Im, C. Mouritsen Luxhoj, A.I. Aviles-Rivero, G.C. Baxter, J.W. MacKay, F.J. Gilbert, Machine learning for workflow applications in screening mammography: systematic review and meta-analysis, Radiology 302 (1) (2022) 88–104.

[6] K. Freeman, J. Geppert, C. Stinton, D. Todkill, S. Johnson, A. Clarke, S. Taylor-Phillips, Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy, BMJ 374 (2021), n1872.

[7] T. Schaffter, D.S.M. Buist, C.I. Lee, Y. Nikulin, D. Ribli, Y. Guan, W. Lotter, Z. Jie, H. Du, S. Wang, J. Feng, M. Feng, H.-E. Kim, F. Albiol, A. Albiol, S. Morrell, Z. Wojna, M.E. Ahsen, U. Asif, A. Jimeno Yepes, S. Yohanandan, S. Rabinovici-Cohen, D. Yi, B. Hoff, T. Yu, E. Chaibub Neto, D.L. Rubin, P. Lindholm, L. R. Margolies, R.B. McBride, J.H. Rothstein, W. Sieh, R. Ben-Ari, S. Harrer, A. Trister, S. Friend, T. Norman, B. Sahiner, F. Strand, J. Guinney, G. Stolovitzky, L. Mackey, J. Cahoon, L.i. Shen, J.H. Sohn, H. Trivedi, Y. Shen, L. Buturovic, J. C. Pereira, J.S. Cardoso, E. Castro, K.T. Kalleberg, O. Pelka, I. Nedjar, K.J. Geras, F. Nensa, E. Goan, S. Koitka, L. Caballero, D.D. Cox, P. Krishnaswamy, G. Pandey, C.M. Friedrich, D. Perrin, C. Fookes, B. Shi, G. Cardoso Negrie, M. Kawczynski, K. Cho, C.S. Khoo, J.Y. Lo, A.G. Sorensen, H. Jung, Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms, J. Am. Med. Assoc. Netw. Open 3 (3) (2020) e200265.

[8] M. Salim, E. Wahlin, K. Dembrower, E. Azavedo, T. Foukakis, Y. Liu, K. Smith, M. Eklund, F. Strand, External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms, J. Am. Med. Assoc. Oncol. 6 (10) (2020) 1581–1588.

[9] N. Houssami, G. Kirkpatrick-Jones, N. Noguchi, C.I. Lee, Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice, Expert. Rev. Med. Dev. 16 (5) (2019) 351–362.

[10] S.M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G.S. Corrado, A. Darzi, M. Etemadi, F. Garcia-Vicente, F. J. Gilbert, M. Halling-Brown, D. Hassabis, S. Jansen, A. Karthikesalingam, C. J. Kelly, D. King, J.R. Ledsam, D. Melnick, H. Mostofi, L. Peng, J.J. Reicher, B. Romera-Paredes, R. Sidebottom, M. Suleyman, D. Tse, K.C. Young, J. De Fauw, S. Shetty, International evaluation of an AI system for breast cancer screening, Nature 577 (7788) (2020) 89–94.

[11] A.D. Lauritzen, A. Rodriguez-Ruiz, M.C. von Euler-Chelpin, E. Lynge, I. Vejborg, M. Nielsen, N. Karssemeijer, M. Lillholm, An artificial intelligence-based mammography screening protocol for breast cancer: outcome and radiologist workload, Radiology 304 (1) (2022) 41–49.

[12] K. Dembrower, E. Wahlin, Y. Liu, M. Salim, K. Smith, P. Lindholm, M. Eklund, F. Strand, Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study, Lancet Digit. Health 2 (9) (2020) e468–e474.

[13] C. Balta, R.W. Bouwman, I. Sechopoulos, M.J.M. Broeders, N. Karssemeijer, R. E. van Engen, W.J.H. Veldkamp, A model observer study using acquired mammographic images of an anthropomorphic breast phantom, Med. Phys. 45 (2) (2018) 655–665.

[14] S. Taylor-Phillips, F. Seedat, G. Kijauskaite, J. Marshall, S. Halligan, C. Hyde, R. Given-Wilson, L. Wilkinson, A.K. Denniston, B. Glocker, P. Garrett, A. Mackie, R. J. Steele, UK national screening committee's approach to reviewing evidence on artificial intelligence in breast cancer screening, Lancet Digit Health 4 (7) (2022) e558–e565.

[15] M.D. Halling-Brown, L.M. Warren, D. Ward, E. Lewis, A. Mackenzie, M.G. Wallis, L. S. Wilkinson, R.M. Given-Wilson, R. McAvinchey, K.C. Young, OPTIMAM mammography image database: A large-scale resource of mammography images and clinical data, Radiol. Artif. Intell. 3 (1) (2021) e200103.

[16] Y.W. Chang, J.K. An, N. Choi, K.H. Ko, K.H. Kim, K. Han, J.K. Ryu, Artificial intelligence for breast cancer screening in mammography (AI-STREAM): A prospective multicenter study design in Korea using AI-based CADe/x, J. Breast Cancer 25 (1) (2022) 57–68.

[17] C. Leibig, M. Brehmer, S. Bunk, D. Byng, K. Pinker, L. Umutlu, Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis, Lancet Digit. Health 4 (7) (2022) e507–e519.

[18] A. Yala, T. Schuster, R. Miles, R. Barzilay, C. Lehman, A deep learning model to triage screening mammograms: A simulation study, Radiology 293 (1) (2019) 38–46.

[19] A. Rodriguez-Ruiz, K. Lang, A. Gubern-Merida, J. Teuwen, M. Broeders, G. Gennaro, P. Clauser, T.H. Helbich, M. Chevalier, T. Mertelmeier, M.G. Wallis, I. Andersson, S. Zackrisson, I. Sechopoulos, R.M. Mann, Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study, Eur. Radiol. 29 (9) (2019) 4825–4832.

[20] K. Lang, M. Dustler, V. Dahlblom, A. Akesson, I. Andersson, S. Zackrisson, Identifying normal mammograms in a large screening population using artificial intelligence, Eur. Radiol. 31 (3) (2021) 1687–1692.

[21] Balta, C., Rodriguez-Ruiz, A., Mieskes, C., Karssemeijer, N. & Heywang-Köbrunner, S. H. Going from double to single reading for screening exams labeled as likely normal by AI: what is the impact? in: Proc. SPIE 11513, 15th Int. Work. Breast Imaging, 2020, pp. 66.

[22] M. Larsen, C.F. Aglen, C.I. Lee, S.R. Hoff, H. Lund-Hanssen, K. Lang, J.F. Nygard, G. Ursin, S. Hofvind, Artificial intelligence evaluation of 122 969 mammography examinations from a population-based screening program, Radiology 303 (3) (2022) 502–511.

[23] A. Graewingholt, P.G. Rossi, Retrospective analysis of the effect on interval cancer rate of adding an artificial intelligence algorithm to the reading process for two-dimensional full-field digital mammography, J. Med. Screen. 28 (3) (2021) 369–371.

[24] D. Byng, B. Strauch, L. Gnas, C. Leibig, O. Stephan, S. Bunk, G. Hecht, AI-based prevention of interval cancers in a national mammography screening program, Eur J Radiol 152 (2022), 110321.

[25] K. Lang, S. Hofvind, A. Rodriguez-Ruiz, I. Andersson, Can artificial intelligence reduce the interval cancer rate in mammography screening? Eur Radiol 31 (8) (2021) 5940–5947.

[26] A.J.T. Wanders, W. Mees, P.A.M. Bun, N. Janssen, A. Rodriguez-Ruiz, M.U. Dalmis, N. Karssemeijer, C.H. van Gils, I. Sechopoulos, R.M. Mann, C.J. van Rooden, Interval cancer detection using a neural network and breast density in women with negative screening mammograms, Radiology 303 (2) (2022) 269–275.

[27] A. Ng, B. Glocker, C. Oberije, G. Fox, N. Sharma, J. James, E. Ambrozay, J. Nash, E. Karpati, S. Kerruish, P. Kecskemethy, Artificial intelligence as supporting reader in breast screening: a novel workflow to preserve quality and reduce workload, J. Breast Imag. (2023).

[28] D. Byng, N. Eisemann, D. Schüler, S. Bunk, C. Leibig, M. Brehmer, S. Elsner, A. Katalinic, Abstract OT3-18-03: The PRAIM study: a prospective multicenter observational study of an integrated artificial intelligence system with live monitoring, Cancer Res. 83 (5_Supplement) (2023). OT3-18-03.

[29] J.L. Raya-Povedano, S. Romero-Martin, E. Elias-Cabot, A. Gubern-Merida, A. Rodriguez-Ruiz, M. Alvarez-Benito, AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: A retrospective evaluation, Radiology 300 (1) (2021) 57–65.

[30] K. Lång, V. Josefsson, A. Larsson, S. Larsson, C. Högberg, H. Sartor, S. Hofvind, I. Andersson, A. Rosso, Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study, Lancet Oncol. 24 (8) (2023 Aug) 936–944.

[31] F. Strand, AI as an independent reader of screening mammograms - results from a prospective study of 55,579 women, in: Presented at RSNA 2022 (S3B-STCE-1). Artificial intelligence in large-scale breast cancer screening (ScreenTrustCAD). Clinicaltrials.gov; NCT04778670. Website visited: May 2nd, 2023.

[32] N. Sharma, A.Y. Ng, J.J. James, G. Khara, E. Ambrozay, C.C. Austin, G. Forrai, G. Fox, B. Glocker, A. Heindl, E. Karpati, T.M. Rijken, V. Venkataraman, J. E. Yearsley, P.D. Kecskemethy, Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms, BMC Cancer 23 (1) (2023) 460.

[33] C.F. de Vries, S.J. Colosimo, R.T. Staff, J.A. Dymiter, J. Yearsley, D. Dinneen, M. Boyle, D.J. Harrison, L.A. Anderson, G. Lip, C. Black, A.D. Murray, K. Wilde, J. D. Blackwood, C. Butterly, J. Zurowski, J. Eilbeck, C. McSkimming, Impact of different mammography systems on artificial intelligence performance in breast cancer screening, Radiol. Artif. Intel. 5 (3) (2023).