Systematic Review of Challenges in Implementing Artificial Intelligence in Breast Cancer Screening Programs: Towards a Framework for Safe Adoption

Serene Goh, Rachel Sze Jen Goh, Bryan Chong, Qin Xiang Ng, Gerald Choon Huat Koh, Kee Yuan Ngiam, Mikael Hartman

Submitted to: Journal of Medical Internet Research
on: June 05, 2024

Table of Contents

Systematic Review of Challenges in Implementing Artificial Intelligence in Breast Cancer Screening Programs: Towards a Framework for Safe Adoption

Serene Goh[1] FRCS; Rachel Sze Jen Goh[2] MBBS; Bryan Chong[2] MBBS; Qin Xiang Ng[3] MBBS; Gerald Choon Huat Koh[3] PHD; Kee Yuan Ngiam[4] FRCS; Mikael Hartman[4] PHD

[1]Yong Loo Lin School of Medicine, National University Singapore Singapore SG
[2]Saw Swee Hock School of Public Health, National University Singapore, Singapore Singapore SG
[3]National University Hospital Singapore Singapore SG

Corresponding Author:
Serene Goh FRCS

Abstract

Background: Artificial intelligence (AI) studies show promise in improving accuracy and efficiency in mammographic screening programs worldwide. However, its adoption in the clinical workflow faces challenges, such as unintended errors, professional training needs, and ethical concerns. Of note, specific frameworks for AI in breast cancer screening are lacking.

Objective: This paper reports a systematic review aiming to assess existing literature and develop a tailored AI governance framework for adoption in breast cancer screening

Methods: Three electronic databases (PubMed, EMBASE, and Medline) were searched using combinations of the key words, “artificial intelligence”, “regulation”, “governance”, “breast cancer” and “screening”. Original studies evaluating AI in breast cancer detection or discussing challenges related to AI implementation in this setting were eligible for review. Findings were narratively synthesized before being mapped directly against the constructs within the Consolidated Framework for Implementation Research (CFIR).

Results: A total of 1240 results were retrieved, and 20 original studies were eventually included in this systematic review. Studies identified challenges in adopting AI in breast screening included reproducibility, evidentiary standards, technology concerns, trust issues, ethical, legal, societal concerns, and post-adoption uncertainty. Mapping these findings against the constructs within the CFIR, we recognize the complex interactions in the development and implementation of a governance framework, across the AI adoption life-cycle in the context of breast cancer screening. Action plans corresponding to the main challenges were included within the framework, aiding in a structured approach to address these issues.

Conclusions: This systematic review identified key themes as well as the barriers and facilitators for AI governance in breast cancer screening. Post-market surveillance is emphasized for continuous monitoring and auditing to ensure the effectiveness and ethical implementation of AI in breast cancer screening. Clinical Trial: na

(JMIR Preprints 05/06/2024:62941)
DOI: https://doi.org/10.2196/preprints.62941

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Systematic Review of Challenges in Implementing Artificial Intelligence in Breast Cancer Screening Programs: Towards a Framework for Safe Adoption

Serene Si Ning Goh[1,2,3], Rachel Sze Jen Goh [3], Bryan Chong[3], Qin Xiang Ng[2], Gerald Choon Huat Koh[2,3,4], Kee Yuan Ngiam [1,2,5], Mikael Hartman[1,2]

[1] National University Hospital, Department of General Surgery, Singapore

[2] National University of Singapore, Saw Swee Hock School of Public Health, Singapore

[3] National University of Singapore, Yong Loo Lin School of Medicine, Singapore

[4] Ministry of Health (MOH) Office of Healthcare Transformation (MOHT), Singapore

[5] Deputy Chief Medical Informatics Officer, National University Hospital, Singapore

Keywords: Breast Cancer, Artificial Intelligence, Screening, Regulations

Corresponding Author

Dr Serene Goh Si Ning

Email: serene_sn_goh@nuhs.edu.sg

National University Hospital, Department of General Surgery, Singapore

Role of funding: There was no funding source for this study. The corresponding author had full access to all data used in the study and had the final responsibility for the decision to submit this article for publication.

Abstract

Background: Artificial intelligence (AI) studies show promise in improving accuracy and efficiency in mammographic screening programs worldwide. However, its adoption in the clinical workflow faces challenges, such as unintended errors, professional training needs, and ethical concerns. Of note, specific frameworks for AI in breast cancer screening are lacking.

Objective: This paper reports a systematic review aiming to assess existing literature and develop a tailored AI governance framework for adoption in breast cancer screening.

Methods

Three electronic databases (PubMed, EMBASE, and Medline) were searched using combinations of the key words, “artificial intelligence”, “regulation”, “governance”, “breast cancer” and “screening”. Original studies evaluating AI in breast cancer detection or discussing challenges related to AI implementation in this setting were eligible for review. Findings were narratively synthesized before being mapped directly against the constructs within the Consolidated Framework for Implementation Research (CFIR).

Results

A total of 1240 results were retrieved, and 20 original studies were eventually included in this systematic review. Studies identified challenges in adopting AI in breast screening included reproducibility, evidentiary standards, technology concerns, trust issues, ethical, legal, societal concerns, and post-adoption uncertainty. Mapping these findings against the constructs within the CFIR, we recognize the complex interactions in the development and implementation of a governance framework, across the AI adoption life-cycle in the context of breast cancer screening. Action plans corresponding to the main challenges were included within the framework, aiding in a structured approach to address these issues.

Conclusion

This systematic review identified key themes as well as the barriers and facilitators for AI governance in breast cancer screening. Post-market surveillance is emphasized for continuous monitoring and auditing to ensure the effectiveness and ethical implementation of AI in breast cancer screening.

Word count: 280 words

Introduction

Breast cancer is the most commonly diagnosed cancer worldwide, with the global prevalence expected to climb in tandem with the ageing population. By 2040, projections show that over 3 million new cases of breast cancer will be diagnosed annually [1]. The increasing global prevalence of breast cancer underscores the urgency of addressing this public health challenge. Many countries worldwide have embraced mammographic screening programs as a vital tool to identify breast cancer at its early stages, significantly reducing the risk of associated mortality [2].

Despite the perceived advantages, there are numerous difficulties in terms of the interpretation of screening mammograms. Firstly, the large number of screening mammograms coupled with the need for double-reading inevitably strains the current radiology workforce [3]. Secondly, the presence of high false positive recall rates on initial screening results in the need for additional procedures and undue anxiety for the patient [4]. Thirdly, approximately 25% of cancers being diagnosed are interval breast cancers, despite regular screening mammograms [5].

Artificial intelligence (AI) presents a solution by automating and streamlining these processes, possibly augmenting efficiency and accuracy. However, the adoption of AI in breast cancer screening is not without challenges. While there are over 20 FDA-approved AI applications for breast imaging, the adoption and utilization of these applications in the clinical settings are widely variable and low overall [6]. Significant barriers to implementation of AI in breast screening include inconsistent performance, limited generalisability of AI algorithms across diverse scenarios, and a lack of confidence among healthcare providers. These challenges underscore the need for well-defined frameworks in the implementation of AI in breast screening.

To date, there are general regulations on the use of AI applications, but none specific to breast cancer screening. This regulatory gap can lead to uncertainties and hesitations in adopting AI technologies in clinical practice. The need for a comprehensive AI governance framework is critical as the medical community considers adopting AI as second readers in screening programs. Hence there is an urgent need to develop a holistic AI governance framework in the face of the ongoing consideration of adopting AI as second readers in screening programs [7-9].

It is known that the implementation process can be affected by a variety of factors, including the characteristics of the institution and its broader environment, as well as the attributes of the individuals delivering the service, who are often practitioners rather than researchers [10,11]. Implementation science serves as a conduit to translate research findings into practical application in real-world settings. To this end, various implementation theories and frameworks have been developed, each tailored to the specific goals of the research [12]. Determinant frameworks, for instance, focus on identifying barriers and facilitators (independent variables) that influence implementation outcomes (dependent variables).

This systematic review employs the Consolidated Framework for Implementation Research (CFIR) [13] as its foundational theoretical framework. The CFIR provides a comprehensive taxonomy of factors that influence implementation across multiple socioecological levels, including community, organizational, and individual levels [13]. This framework is particularly relevant for implementing AI in breast cancer screening, as it helps identify and address the multifaceted barriers and facilitators involved in this complex process. Through this review, we also aim to appraise the available literature and propose a working framework for the safe adoption of AI imaging in breast cancer screening.

Methods

Search Strategy and selection criteria

This systematic review was registered on the international prospective register of systematic reviews (PROSPERO), registration number 531852 and is structured in line with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) checklist [14]. For this systematic review, we searched literature from inception of the databases to 30 January 2024, to identify articles with guidance for the development of a governance framework for adoption of AI in breast cancer screening. The literature search was restricted to articles between 1 January 2000 to 30 January 2024. The databases searched include PubMed, EMBASE and Medline. Keywords employed in the search include “artificial intelligence”, “regulation”, “governance”, “breast cancer” and “screening”. The full search terms and strategy are displayed in Supplementary Material 1.

Two reviewers (RSJG, BC) independently evaluated the articles both in the title abstract and full-text stage based on a priori inclusion and exclusion criteria, and a third independent reviewer (SG) was consulted when a consensus could not be reached. Inclusion criterion was studies that provide guidance on evaluating AI in breast cancer detection. Exclusion criterion was studies unrelated to AI in breast cancer detection or non-English articles or abstracts.

Data analysis and synthesis

Data collection was conducted in a blinded manner by two independent and blinded reviewers (RSJG, BC) in a pre-determined data collection form. Variables collected include title, author, year, country, objective, methods, and reporting of key challenges and solutions. The findings were synthesized using a narrative synthesis approach. Initially, a preliminary synthesis was conducted through thematic analysis, which involved searching for relevant studies, listing them, and presenting the results in tabular form. Subsequently, the results were discussed and organized into themes. Finally, the included studies were summarized in a narrative synthesis within the CFIR framework [13], with all authors reaching a consensus. This framework comprised the following constructs: intervention characteristics (relative advantage, adaptability, complexity, design factors), outer setting (patient needs, resources, policy, and incentives), inner setting (infrastructure, networks, implementation climate), and individual characteristics (knowledge, beliefs, self-efficacy).

Quality Assessment

The relevance and quality of the selected studies were assessed independently by at least two reviewers in reference to quality domains adapted from Batini, Cappiello, Francalanci, and Maurino (2009) [15] and (Bano & Zowghi, 2015) [16]:

- Accuracy: The objectives of the study are clearly stated with the data collection methodology adequately described. Important statements should be supported with references.
- Consistency: The design of the study is appropriate for the research objectives. The study's research questions are answered or the research objective is attained.
- Completeness: The study's research approach is described in sufficient detail.
- Timeliness: The study was published in the past ten years.
- Relevance: An additional domain was included to ensure that the study was relevant and had substantial discussion evaluating the use of AI in breast cancer screening.

The studies were assessed in the respective domains on a scale of one to five, with one representing the minimal relevance in the domain and five representing high relevance in the domain. The quality of the studies were independently assessed by two independent reviewers (RSJG, BC), with a third author (SG) resolving any disagreements.

Results

A total of 1240 abstracts were retrieved from the databases for the initial sieve. After 28 duplicates were removed, a priori inclusion and exclusion criteria were then applied at two stages: the title abstract sieve, and the full-text sieve. A total of 20 articles [17-35] were included in this systematic review, with the abstraction process shown in Figure 1. Their key characteristics and findings are summarized in Table 1. In terms of breast cancer screening modalities, 19 articles were on breast screening in general or mammograms in breast screening, whilst one article was on ultrasound which is relevant in the context of supplemental imaging for women with dense breasts. Majority of studies were from USA (N=5), United Kingdom (N=4), Australia (N=2), Saudi Arabia (N=1), Pakistan (N=1), France (N=1) and Korea (N=1); 4 studies were global studies. The year of publication of articles ranged from 2019-2023. All 20 studies were rated to be moderate to high quality based on the quality assessment and met our inclusion criteria; a full quality assessment is appended in Supplementary Material 2.

Through iterative discussions, eight themes were identified based on reading of the studies. They encompass the challenges in the adoption of AI in breast screening were identified: reproducibility, evidentiary standards, technology concerns, trust issues, ethical concerns, legal concerns, societal concerns and uncertainty post-adoption of AI in breast cancer screening. Dataset and validation limitations have been described by the majority of articles [19,25,29]. Common concerns were the quality of the dataset and the lack of transparency of the data being used for AI development [29]. Ascertainment bias may arise when cancer detection is based on decision by human readers instead of alternatives such as histopathological diagnosis, impacting the reliability of ground truths used for training or evaluation of cancer detection algorithms. Sufyan et al. [31] described the lack of and inconsistency in the annotation of data used to train an AI model which can result in unreliable predictions. When an AI software is trained on biased data (for example reflecting only certain race or age group) or has inherent algorithm bias, it inevitably results in discriminatory outcomes. For AI systems developed based on mammograms which have underwent resizing, augmentation or specific segmentation process, they may encounter issues with lesion classification performance. With regards to AI in ultrasound screening, J Kim et al. [33] reported concerns about high interobserver variability during obtaining and interpretation of the images may result in diagnostic inaccuracies and thus management discrepancies. Additionally, Gastounioti et al. [21] suggested that the large variability could affect the perceived clinical applicability of AI-generated risk assessment based on mammographic assessments, emphasizing that

reproducibility, generalizability and interpretability were fundamental principles to encourage the translation of AI into clinical practice.

In terms of evidentiary standards, several studies reported the quality of supporting evidence to be a primary concern in determining the utility of AI in the clinical setting [23,27]. Evidentiary standards have been largely limited by the use of data-enriched datasets and small sample sizes [31,32]. Cancer-enriched data sets contain more true positive cases than in a conventional screening setting. The risk of false positives is likely to increase in such datasets, translating to unnecessary investigations, biopsies and anxiety. Limited, single-center studies or studies that lack clinical diversity in terms of patients of various ages, breast density, and breast cancer risks may result in poor generalizability [31,32]. The paucity of clinical validation and patient-centric outcomes were also highlighted [18,35]. Evaluation processes have been mainly focused on AI's performance rather than actual clinical outcomes for patients and health systems. Moreover, cancer detection by AI has not been shown to translate to improved health. Taylor-Phillips et al. [28] suggested the use of clinically significant and relevant outcomes (i.e. interval breast cancers) to evaluate the overall effect of AI, given the potential downstream effects that false positives on screening many have on the allocation of healthcare resources. The absence of standardized performance metrics across studies makes it challenging to compare the effectiveness of different AI models [33].

With regard to the technological concerns and requirements, Thomassin-Naggara et al. [35] highlighted the need for large storage capacity for massive data whilst Lamb et al. [22] the need to ensure compatibility of the AI system with local practice techniques and equipment. Seamless integration of AI systems with existing healthcare information systems, such as picture archiving and communication systems [43] and electronic health records is critical for effective collaboration and data sharing. AI algorithms, especially deep learning models, often require significant computational power. The technical expertise to develop and maintain IT infrastructure [24] is also required to support AI systems and ensure scalability in breast cancer screening programs. Additionally, radiologists will require training to understand the appropriate use of various tools and their limitations [23,25,37]. In terms of AI user-interface, through qualitative interviews, Hendrix et al. [24] found that 26–33% of radiologists were deterred if the AI features did not align with their preferences, highlighting the role for increased collaboration between radiologists and technical professionals to bridge this gap.

Pertaining to trust issues, Lamb et al. [22] described the variability in the level of trust in AI by radiologists, non-radiologist clinicians, and patients. Similarly, Hendrix et al. [24] reported that physicians remain wary of the use of AI for unsupervised or partially supervised image interpretation. Carter et al. [34] and Bahel et al. [37] postulated that this is related to the tendency of AI to recommend individualized decisions that are not explainable, instead of providing general recommendations at a population level such as in a conventional screening setting. The term “black box” refers to an AI system whose internal workings are not transparent or easily understandable [32]. Intellectual property clauses protecting proprietary sources fuel the lack of transparency by limiting the distribution/sharing of the AI source code and architecture [44]. With limited human involvement in the decision-making process and a lack of rationale, there is significant scepticism and reduced trust on the decisions made with the help of AI [23,27,47].

Ethical concerns do exist as well. Biases present in training data can be perpetuated by AI algorithms [17,21,32], leading to disparities in breast cancer screening accuracy across different demographic groups. Outcomes based on pre-existing inequalities could be compounded by the skewed data being fed into the algorithm, creating negative reinforcement and further worsening the inequality [25]. However, this can potentially be mitigated by carefully selecting training data and strategically developing models and frameworks keeping these underrepresented demographics in mind [21]. K Badal et al. [30] highlighted the specialized resources and expertise required for AI adoption themselves pose barriers for entry into systems serving the most disadvantaged populations. In addition, Al Kuwaiti A et al. [18] discussed the concept of lack of transparency and explainability of AI, which raises ethical concerns about accountability. Lamb et al. [22] and Carter et al. [34] raised the presence of conflict of interests as an ethical issue, especially between clinicians and developers which may influence the use of AI systems. K Badal et al. [30] and Carter et al. [34] both emphasized the need for shared decision making, where women undergoing screening should have the autonomy to choose whether to incorporate AI-enabled screening into their healthcare.

With regard to legal challenges, data breaches in the context of AI are particularly concerning, and they refer to incidents where unauthorized individuals or entities gain access to sensitive and confidential information processed or stored by AI systems. These breaches can have significant consequences, ranging from privacy violations to identity theft and other forms of cybercrime. Another legal challenge is the determination of legal

liability if AI systems in breast screening make errors leading to misdiagnosis or patient harm [25]. Determining responsibility for algorithmic errors, whether it lies with the developers, healthcare providers, or both, raises complex legal questions. Potnis et al. [32] stated that the FDAs' role is for the maintenance of a minimum threshold for AI product approval and suggested that interested parties have responsibilities over how AI is being adopted. Carter et al. [21] described a regulatory vacuum for AI-related technologies. The absence of clear regulations can create uncertainty for stakeholders about the legal boundaries associated with AI adoption.

Concerns were also raised regarding the impact of AI on professional development and society. Fears of job displacement and dependency among radiologists may occur if AI takes over the role of radiologists in interpreting normal mammograms [27]. Likewise, Carter et al. [21] and Hickman et al. [25] proposed that if radiologists excessively rely on AI for interpreting normal mammograms, they might experience a reduction in skills or lack familiarity with normal images, resulting in the oversight of important clinical nuances. Other potential biases to consider include automation biases, where readers are greatly influenced by AI generated decisions and over commit to false positives or omit other abnormalities. Additionally, there is a potential for the "anchoring effect" – in which once markers for potential malignancy are placed in the image, it has the tendency to influence their decision making [19]. There are also concerns regarding an eventual environment with reduced creativity and emotions if AI replaces radiologists [18]. Radiologists may resist the adoption of AI in breast screening due to concerns about the reliability of the technology, fear of job loss, or scepticisms about the ability of AI to outperform human expertise [22].

There was also the sense of uncertainty post-adoption as AI models may encounter diverse and evolving conditions that were not fully represented in the training data. Logan et al. [17] described the potential of unintended consequences of AI which may affect stakeholders such as patients, radiologists, and institutes. Adverse events may include misdiagnoses, false positives, or false negatives to patients. Alongside these identified themes, proposed solutions were illustrated in Figure 2.

Applying the CFIR framework, under the domain of intervention characteristics, AI in breast cancer screening shows a clear relative advantage by improving the accuracy and efficiency of mammographic interpretation. This advantage is bolstered by the adaptability of AI tools, which can be tailored to fit various clinical settings

and diverse populations. However, the complexity of integrating AI into existing clinical workflows presents a significant challenge, as it requires substantial technical expertise and extensive training for healthcare providers. Furthermore, the design quality and packaging of AI tools are crucial, necessitating user-friendly interfaces that align with radiologists' preferences to facilitate adoption.

In the outer setting domain, addressing patient needs and resources, AI has the potential to reduce false positives and associated patient anxiety, improving the overall patient experience in breast cancer screening. However, the evolution of external policy and incentives is less defined, as current regulatory frameworks must adapt to support the integration of AI in clinical practice. Regulatory clarity and support can help alleviate uncertainties and foster a conducive environment for AI adoption in breast cancer screening programs.

Within the inner setting domain, several studies highlighted that the successful implementation of AI requires robust digital infrastructure to handle the large volumes of data generated and processed by AI systems. Effective networks and communications are essential for seamless integration, necessitating collaboration among radiologists, IT professionals, and AI developers. Additionally, the implementation climate, including organizational readiness for change and the perceived value of AI tools, significantly impacts the adoption process. Institutions must be prepared and supportive of these technological advancements to ensure successful implementation.

Last but not least, with regard to the characteristics of individuals involved in the implementation process, knowledge and beliefs about AI interventions play a significant role [47], as trust issues regarding the accuracy and transparency of AI systems need to be addressed. Building self-efficacy through comprehensive training programs helps to enhance radiologists’ confidence in using AI tools and to mitigate fears of job displacement and deskilling. The process of implementing AI involves careful planning, including structured plans, pilot testing, and phased rollouts. Engaging key stakeholders—such as radiologists, patients, and regulatory bodies—is vital to address concerns and secure buy-in. Continuous monitoring and iterative adjustments based on feedback are necessary to ensure effective execution, while ongoing evaluation of AI performance and clinical outcomes is critical for sustained improvement and success.

Discussion

Through this review, it is evident that several critical aspects influence the implementation of AI in breast cancer screening. A significant concern across numerous studies is the reproducibility of AI algorithms, which hinges on the quality and transparency of the datasets used for their development. Logan et al. [17] suggested the improvement of AI datasets by achieving findability, accessibility, interoperability and reusability (FAIR principles [36]) of the data. Adopting best practices such as open-sourcing code, providing detailed documentation, and promoting transparency in dataset setups can contribute to improving reproducibility.

In research and algorithm development, to reduce ascertainment bias, efforts should be made to use gold standard reference data whenever possible such as histopathology results. Thomassin-Naggara et al. [22] described the development of a universal annotation tool. Training and calibrating annotators' judgments can reduce inter-annotator variability as well. Data quality control measures, such as regular reviews and feedback loops, can help identify and correct inconsistencies that may arise. Additionally, validation of the AI algorithm on diverse and representative datasets, especially local datasets, enhanced the generalizability of cancer detection algorithms and their performance in real-world scenarios [21,37]. As alluded to by Hickman et al. [25], a reliable algorithm with clear, consistent and reproducible results, with minimal ambiguity in decision-making, is integral to improving confidence in AI systems.

The quality of supporting evidence for AI's clinical use is another major concern. Cushnan et al. [20] alluded to prospective randomised controlled trials as the gold standards for comparing the performance of AI to human readers. However, due to resource constraints and the rapid progression of technology where new AI systems will be released from time to time, an alternative method such as virtual clinical trials evaluates whether readers at arbitration would recall an interval cancer identified by AI but not by the human reader. Le et al. [19] and Taylor Phillips et al. [28] suggested that a stepwise approach where retrospective evaluation of the AI software with a large representative dataset is employed first for benchmarking, followed by prospective evaluation in the clinical setting. Large, consecutive retrospective studies may also be beneficial in providing useful information on population subgroups which may be underpowered in prospective studies [28]. Nijnatten et al. [26] suggested a similar framework, with the preference for retrospective trials as the time-consuming and resource-intensive nature of prospective trials could hinder the adoption of AI in the clinical setting.

A dedicated platform to build large validation data sets representative of the screening population will be optimal [20.35], enabling the comparison of different AI systems with one another and radiologist-read mammograms, the current standard of care [28]. Enriched datasets can also prove useful in facilitating the analysis of AI use in minority and underrepresented populations [20], improving their overall accuracy. Simulation models can shed light on how an AI system can enhance breast screening outcomes such as morbidity and mortality in the long term [17]. To ensure AI solutions have net benefits to patients, their development and design choices should abide by the IDEAL framework [34,38]. Furthermore, consistent metrics are crucial for establishing benchmarks and evaluating progress in this field. Pre-clinical reporting guidelines such as TRIPOD-AI [39], clinical reporting guidelines such as STARD-AI [40], DECIDE-AI [41], CONSORT-AI [42], SPIRIT-AI [43] can contribute to the generation of standardized and quality evidence in the field. Studies should also improve focus on the clinical utility of AI in terms of its impact on treatment decisions, patient outcomes and healthcare systems.

Several prerequisites for maintenance of AI systems include digital infrastructure available, technical skills to manage vast data and sufficient computational power [22]. Maintaining a strong relationship between domain experts such as breast radiologists and technical experts including data scientists and engineers is essential to ensure that AI solutions remain clinically relevant, user-friendly, and aligned with the ethical and practical considerations of the healthcare domain. This is especially so given Hendrix et al. [24] found that radiologists' endorsement of AI usage may depend on the distinct characteristics and capabilities it possesses.

As highlighted in the CFIR model, trust in AI systems among radiologists, other healthcare providers, and patients is a pivotal factor for successful adoption. A method to improve the trust in AI as suggested by Al Kuwaiti A et al. [18] was to provide visual feedback to the user regarding metrics used to obtain breast cancer prediction. In the context of breast screening, superimposed heatmaps over suspicious areas of the mammogram or ultrasound could be an option. Carter et al. [34] underlined the importance of setting clear expectations to stakeholders regarding explainability of AI processes. The mandatory involvement of human radiologist on decisions regarding final screening outcomes may reduce the level of public distrust in this relatively new field of AI advancement. AI systems can assist in the screening process, providing additional insights, but the final decision and interpretation should involve human judgment. Furthermore, public accountability and public engagement regarding AI enabled healthcare are required to sustain peoples' trust in

healthcare systems.

As aforementioned, ethical concerns also arise from biases in training data, which can perpetuate disparities in screening accuracy across different demographic groups. K Badal et al. [30] suggested strategies to reduce health disparity such as training AI on balanced datasets representing disadvantages populations and prioritizing the development of AI tools for hospitals serving the disadvantaged. However ethical considerations and the actual accessibility of interventions should be considered. In face of conflict of interest between clinicians and developers, Lamb et al. [20] suggested recusal of members with a significant conflict of interest from the decision-making process of AI adoption. Finally, both K Badal et al. [30] and Carter et al. [34] suggested that women should be given the autonomy to decide regarding the use of AI-based screening. K Badal et al. [30] aptly highlighted that this would require patient's and practitioners to understand the pros and cons of AI-based screening prior to decision-making, and this is only possible if the AI tools are designed to improve patient understanding and are more patient-centric [45]. A method proposed by Birch et al. [46] was the generation of a continuous risk score instead of a fixed score so that patients can make informed decision based on their risk preference.

Collaboration between AI developers, healthcare providers, legal experts, and regulatory bodies is also essential to navigate the complex legal landscape surrounding the use of AI in healthcare. Carter et al. [21] discussed the adaption of existing frameworks or development new purpose-built regulatory framework to safeguard patient privacy and account for potential liabilities. To address accountability, comprehensive frameworks that outline the responsibilities of stakeholders at each stage of AI deployment will be essential to outline liability frameworks. Ensuring compliance with data protection laws and avoiding data breaches can safeguard patient privacy.

Hickman et al. [25] has also suggested a role for further guidance on accountability of the companies who developed the algorithm. "Accidents" and "near misses" arising from the use of AI, should be reviewed and reported to the respective companies to facilitate algorithm improvement to mitigate future risks of recurrence.

Nevertheless, the public health impact of AI adoption is still ambiguous as studies lack focus on patient-centric outcomes and cost effectiveness analysis. Investing in training programs for radiologists, technologists, and

administrators to familiarize them with AI tools and AI workflow integration may mitigate the concerns of deskilling [22]. Carter et al. [21] suggested that radiologists should be trained to avoid automation bias as well. Although AI adoption may result in skewed decision-making from excessive dependence on AI, it is important to acknowledge that the current breast screening landscape is not without inherent issues such as cognitive bias too. In terms of fostering trust and addressing concerns within the broader community, public engagement and transparent communication are also important. Future research on AI in breast cancer detection should prioritize examining clinical outcomes.

Post-market surveillance for AI in breast screening is also vital for ensuring ongoing safety, efficacy, and relevance of the technology in real-world healthcare settings. It enables prompt intervention to rectify issues and mitigate potential harm to patients. Logan et al. [17] suggested leveraging on reporting structures used for quality assurance to ensure adherence to safety and performance standards. Post-market surveillance also allows improvements based on user feedback on the practical usability and limitations of AI systems. K Badal et al. [27] emphasized the importance of designing AI algorithms with the intent of regular evaluation, learning, and improvement. This is crucial due to the potential for unexpected errors in AI deployment and the need to integrate new scientific knowledge for patient

Using Singapore and our own experience as a case study, to facilitate development and implementation of policies related to artificial intelligence in breast cancer screening, the three key components of the actors, context, and processes are described sequentially in Figure 3. Further refinement will require a collaborative and iterative process, so that policymakers can tailor their strategies to ensure that the AI adoption framework is responsive to the unique challenges and opportunities in each country. Its envisioned use would empower stakeholders to navigate the complex policy landscape and harness the potential of AI to improve breast cancer detection and care.

Strengths and limitations

This systematic review lays the groundwork for developing a comprehensive AI governance framework for breast cancer screening. Given the burgeoning interest in this field, our methodology was both timely and appropriate for gathering the necessary information. The identified themes are interdependent, highlighting the real-world complexities of AI implementation. For example, issues with reproducibility directly impact

evidentiary standards, while ethical concerns about data breaches can lead to significant legal ramifications. The rapidly evolving nature of AI necessitates regular reviews of new publications to remain current with the latest developments. Future efforts should incorporate additional methodologies such as focused group discussions with stakeholders, real-world case study analyses, and international collaborations. These approaches can provide deeper insights and more robust contributions to the development of an effective AI governance framework. However, this review also uncovered significant imbalances in the available literature across different domains. While there is ample discussion on reproducibility, evidentiary standards, and technological concerns, there is a relative paucity of studies addressing legal issues and post-adoption uncertainty. This gap underscores the need for future research to delve more deeply into the legal and ethical dimensions of AI use in breast cancer screening. Addressing these gaps will be crucial for creating a holistic and effective governance framework that can be widely adopted in clinical practice.

Conclusion

This systematic review has identified and mapped the key themes of AI governance in breast cancer screening against the CFIR framework. These themes encompass crucial domains such as data reproducibility, high-quality evidentiary standards, technological readiness of radiologists, and the explainability of AI algorithms to build trust among medical professionals and patients. Additionally, the review highlights the importance of ethical, legal, and societal considerations, as well as the necessity for post-market surveillance to ensure continuous monitoring and auditing of AI systems. To maximize the benefits of AI in breast cancer screening, it is essential to design algorithms that prioritize regular evaluation and improvement. This approach will facilitate the integration of new scientific knowledge and ultimately enhance patient outcomes. A comprehensive AI governance framework, informed by these themes, will be instrumental in guiding the safe and effective adoption of AI technologies in clinical practice.

Table, Figures and Supplementary Materials

List of abbreviations

Artificial Intelligence – AI

Availability of Data and Materials

The data analyzed during the current study are available from the corresponding author on reasonable request. Additionally, all included studies were sourced from publicly available databases such as PubMed, Embase, and Web of Science. Detailed search strategies and inclusion criteria are provided in the supplementary materials.

Funding

Serene Goh Si Ning has received funding from NUHS Clinician Scientist Program 2.0 (NCSP 2.0) under Department of Surgery, National University Hospital, and Associate Professor Mikael Hartman has received funding from Breast Cancer Screening Prevention Programme (NUHSRO/2020/121/BCSPP/LOA) under Yong Loo Lin School of Medicine, National University of Singapore. However, these funding did not influence the design, conduct, or interpretation of the study presented in this article.

Competing interest declaration

The authors declare that they have no competing interests.

Author's contribution

SSN Goh: Conceptualize the study, Literature review, manuscript writing, Designing of Figures, Tables

RSJ Goh; Literature review, Data cleaning, manuscript writing

B. Chong; Literature review, Data cleaning, manuscript writing

QX. Ng; Literature review, manuscript writing

GC. Huat. Koh; Manuscript writing and guidance

Ngiam. KY: Manuscript writing and guidance

M. Hartman: Manuscript writing and guidance

Ethics approval and consent to participate, consent for publication

Not applicable in this study design

Acknowledgement

Not applicable

References

1. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. Dec 2022;66:15-23. doi:10.1016/j.breast.2022.08.010
2. Dibden A, Offman J, Duffy SW, Gabe R. Worldwide Review and Meta-Analysis of Cohort Studies Measuring the Effect of Mammography Screening Programmes on Incidence-Based Breast Cancer Mortality. Cancers (Basel). Apr 15 2020;12(4)doi:10.3390/cancers12040976
3. Tay YX, Kothan S, Kada S, Cai S, Lai CWK. Challenges and optimization strategies in medical imaging service delivery during COVID-19. World J Radiol. May 28 2021;13(5):102-121. doi:10.4329/wjr.v13.i5.102
4. Ho T-QH, Bissell MCS, Kerlikowske K, et al. Cumulative Probability of False-Positive Results After 10 Years of Screening With Digital Breast Tomosynthesis vs Digital Mammography. JAMA Network Open. 2022;5(3):e222440-e222440. doi:10.1001/jamanetworkopen.2022.2440
5. Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. Int J Environ Res Public Health. Apr 23 2017;14(4)doi:10.3390/ijerph14040452
6. Taylor CR, Monga N, Johnson C, Hawley JR, Patel M. Artificial Intelligence Applications in Breast Imaging: Current Status and Future Directions. Diagnostics. 2023;13(12):2041.
7. Sharma N, Ng AY, James JJ, et al. Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms. BMC Cancer. 2023/05/19 2023;23(1):460. doi:10.1186/s12885-023-10890-7
8. Ng AY, Glocker B, Oberije C, et al. Artificial Intelligence as Supporting Reader in Breast Screening: A Novel Workflow to Preserve Quality and Reduce Workload. Journal of Breast Imaging. 2023;5(3):267-276. doi:10.1093/jbi/wbad010
9. Ng AY, Oberije CJG, Ambrózay É, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nature Medicine. 2023/12/01 2023;29(12):3044-3049. doi:10.1038/s41591-023-02625-9
10. Theobald S, Brandes N, Gyapong M, El-Saharty S, Proctor E, Diaz T, Wanji S, Elloker S, Raven J, Elsey H, Bharal S, Pelletier D, Peters DH. Implementation research: new imperatives and opportunities in global health. Lancet. 2018 Nov 17;392(10160):2214-2228. doi: 10.1016/S0140-6736(18)32205-0.
11. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ. 2013 Nov 20;347:f6753. doi: 10.1136/bmj.f6753.
12. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci. 2015 Apr 21;10:53. doi: 10.1186/s13012-015-0242-0.
13. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009 Aug 7;4:50. doi: 10.1186/1748-5908-4-50.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
15. Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM computing surveys (CSUR). 2009;41(3):1-52.
16. Bano M, Zowghi D. A systematic review on the relationship between user involvement and system success. Information and software technology. 2015;58:148-169.
17. Logan J, Kennedy PJ, Catchpoole D. A review of the machine learning datasets in mammography, their adherence to the FAIR principles and the outlook for the future. Scientific Data. 2023/09/08 2023;10(1):595. doi:10.1038/s41597-023-02430-6
18. Al Kuwaiti A, Nazer K, Al-Reedy A, et al. A Review of the Role of Artificial Intelligence in Healthcare. Journal of Personalized Medicine. 2023;13(6):951.
19. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol. May 2019;74(5):357-366. doi:10.1016/j.crad.2019.02.006
20. Cushnan D, Young KC, Ward D, et al. Lessons learned from independent external validation of an AI tool to detect breast cancer using a representative UK data set. Br J Radiol. Mar 2023;96(1143):20211104. doi:10.1259/bjr.20211104
21. Gastounioti A, Desai S, Ahluwalia VS, Conant EF, Kontos D. Artificial intelligence in mammographic phenotyping of breast cancer risk: a narrative review. Breast Cancer Research. 2022/02/20 2022;24(1):14. doi:10.1186/s13058-022-01509-z
22. Lamb LR, Lehman CD, Gastounioti A, Conant EF, Bahl M. Artificial Intelligence (AI) for Screening Mammography, From the AJR Special Series on AI Applications. AJR Am J Roentgenol. Sep

2022;219(3):369-380. doi:10.2214/ajr.21.27071
23. Hendrix N, Hauber B, Lee CI, Bansal A, Veenstra DL. Artificial intelligence in breast cancer screening: primary care provider preferences. J Am Med Inform Assoc. Jun 12 2021;28(6):1117-1124. doi:10.1093/jamia/ocaa292
24. Hendrix N, Lowry KP, Elmore JG, et al. Radiologist Preferences for Artificial Intelligence-Based Decision Support During Screening Mammography Interpretation. J Am Coll Radiol. Oct 2022;19(10):1098-1110. doi:10.1016/j.jacr.2022.06.019
25. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. British Journal of Cancer. 2021/07/01 2021;125(1):15-22. doi:10.1038/s41416-021-01333-w
26. van Nijnatten TJA, Payne NR, Hickman SE, Ashrafian H, Gilbert FJ. Overview of trials on artificial intelligence algorithms in breast cancer screening - A roadmap for international evaluation and implementation. Eur J Radiol. Oct 2023;167:111087. doi:10.1016/j.ejrad.2023.111087
27. Singh A, Paruthy SB, Belsariya V, et al. Revolutionizing Breast Healthcare: Harnessing the Role of Artificial Intelligence. Cureus. Dec 2023;15(12):e50203. doi:10.7759/cureus.50203
28. Taylor-Phillips S, Seedat F, Kijauskaite G, et al. UK national screening committee's approach to reviewing evidence on artificial intelligence in breast cancer screening. The Lancet Digital Health. 2022;4(7):e558-e565.
29. Tran WT, Sadeghi-Naini A, Lu FI, et al. Computational Radiology in Breast Cancer Screening and Diagnosis Using Artificial Intelligence. Can Assoc Radiol J. Feb 2021;72(1):98-108. doi:10.1177/0846537120949974
30. Badal K, Lee CM, Esserman LJ. Guiding principles for the responsible development of artificial intelligence tools for healthcare. Communications Medicine. 2023/04/01 2023;3(1):47. doi:10.1038/s43856-023-00279-9
31. Sufyan M, Shokat Z, Ashfaq UA. Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective. Comput Biol Med. Oct 2023;165:107356. doi:10.1016/j.compbiomed.2023.107356
32. Potnis KC, Ross JS, Aneja S, Gross CP, Richman IB. Artificial Intelligence in Breast Cancer Screening: Evaluation of FDA Device Regulation and Future Recommendations. JAMA Intern Med. Dec 1 2022;182(12):1306-1312. doi:10.1001/jamainternmed.2022.4969
33. Kim J, Kim HJ, Kim C, Kim WH. Artificial intelligence in breast ultrasonography. Ultrasonography. Apr 2021;40(2):183-190. doi:10.14366/usg.20117
34. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast. Feb 2020;49:25-32. doi:10.1016/j.breast.2019.10.001
35. Thomassin-Naggara I, Balleyguier C, Ceugnart L, et al. Artificial intelligence and breast screening: French Radiology Community position paper. Diagn Interv Imaging. Oct 2019;100(10):553-566. doi:10.1016/j.diii.2019.08.005
36. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016/03/15 2016;3(1):160018. doi:10.1038/sdata.2016.18
37. Bahl M. Artificial Intelligence: A Primer for Breast Imaging Radiologists. J Breast Imaging. Aug 2020;2(4):304-314. doi:10.1093/jbi/wbaa033
38. The IDEAL Framework. Accessed 8 March 2024, 2024. https://www.ideal-collaboration.net/the-ideal-framework/
39. Gary SC, Paula D, Constanza LAN, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008. doi:10.1136/bmjopen-2020-048008
40. Viknesh S, Hutan A, Robert MG, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11(6):e047709. doi:10.1136/bmjopen-2020-047709
41. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine. 2022/05/01 2022;28(5):924-933. doi:10.1038/s41591-022-01772-9
42. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine. 2020/09/01 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x
43. Ibrahim H, Liu X, Rivera SC, Moher D, Chan AW, Sydes MR, Calvert MJ, Denniston AK. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI

guidelines. Trials. 2021 Jan 6;22(1):11. doi: 10.1186/s13063-020-04951-6.
44. Strickland NH. PACS (picture archiving and communication systems): filmless radiology. Archives of disease in childhood. 2000/07// 2000;83(1):82-86. doi:10.1136/adc.83.1.82
45. Gilbert FJ, Smye SW, Schönlieb CB. Artificial intelligence in clinical imaging: a health system approach. Clin Radiol. Jan 2020;75(1):3-6. doi:10.1016/j.crad.2019.09.122
46. Birch J, Creel KA, Jha AK, Plutynski A. Clinical decisions using AI must consider patient values. Nature Medicine. 2022/02/01 2022;28(2):229-232. doi:10.1038/s41591-021-01624-y
47. Gama, F., Tyskbo, D., Nygren, J., Barlow, J., Reed, J., & Svedberg, P. (2022). Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review. Journal of medical Internet research, 24(1), e32215. https://doi.org/10.2196/32215

Supplementary Files