Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240480

Enabling AI in Radiology: Evaluation of an AI Deployment Process

Line SILSAND[a,1], Mari KANNELØNNING[b], Gro-Hilde SEVERINSEN[b] and Gunnar ELLINGSEN[a,b,]
[a] *The Arctic University of Norway, Norway*
[b] *Norwegian Centre for E-health Research, Norway*
ORCiD ID: Line Silsand https://orcid.org/0000-0003-0971-773X

Abstract. Artificial intelligence (AI) is expected to transform healthcare systems and make them more sustainable. Despite the increased availability of AI tools for disease detection, evidence of their impact on healthcare organisations and patient care remains limited. Drawing on previous research underscoring the need for comprehensive evaluations of real-world AI deployments, this paper explores the challenges and opportunities encountered while procuring and implementing AI solutions for radiology. The paper aims to contribute to a better understanding of the complexities surrounding AI deployments in real-world clinical settings through a process evaluation study.

Keywords. Artificial intelligence, radiology, deployment, process evaluation

1. Introduction

During the last six years, a multitude of AI tools designed to assist radiologists in image diagnostics have become commercially available following regulatory approval [1-3]. However, this approval is typically based on retrospective evidence showing the diagnostic performance of an AI tool trained on certain datasets [2,4,5]. With such processes as the primary validation approach, the performance of AI tools in real-world clinical settings often remains hidden and unverified [3,4]. This less scrutinised part of how AI tools function in actual use underscores the need for a more nuanced analytical framework to evaluate their performance as they enter clinical settings [4,5]. There is a need for approaches in terms of validation and evaluation that take a broader perspective beyond the existing retrospective evidence of diagnostic performance. Such evaluations are necessary to enable decision-makers to make informed decisions regarding AI deployment in healthcare services. Thus, this paper aims to show how process evaluation can supplement existing retrospective validation and bring forth a broader perspective by identifying barriers or challenges necessary to address during early phases of complex interventions in clinical practice (i.e. AI deployments).

Drawing on a process evaluation study of an AI procurement and implementation process at a large Norwegian hospital trust [6], this paper seeks to answer the following

[1] Corresponding Author: Line Silsand, UiT - The Arctic University of Norway; E-mail: line.silsand@uit.no.

research question: *What kind of challenges confront the procurement and implementation of AI for radiology practices, and how can they be managed?* The process evaluation was guided by the Non-Adoption or abandonment, Scale-up, spread and sustainability framework (NASSS) [7]. Additionally, the Information Infrastructure framework [8] is applied to define and discuss the socio-technical factors influencing the deployment.

2. Methods

Since 2020, we have carried out a process evaluation of a procurement process and implementation of commercialised AI algorithms in a Norwegian hospital trust. Our research includes 59 semi-structured interviews with hospital managers and healthcare personnel (radiologists, radiographers, orthopaedists and emergency room physicians). We have conducted approximately 50 hours of observations during the procurement phase and attended status meetings with the trust's project team every third week (from autumn 2021). During the latter, we discussed the project status and our findings as the deployment process proceeded. Finally, we participated in various meetings related to the practical preparations of the implementation process (e.g., workshops mapping out existing and future radiology workflows). Our data analysis involved systematic reading and thematic coding of interviews guided by the NASSS framework [7] supplemented by the meeting fieldnotes. Altogether, the data analysis employed a hermeneutic approach, considering all collected data to get a balanced picture of the process [9]. The data is presented as an empirical narrative illustrating the multifaceted reality and the imperative for action in procuring and implementing CE-marked AI solutions in a public hospital trust.

3. Results

Procurement phase: Insufficient documentation and limited evidence

The Norwegian hospital trust studied perceived that procuring commercially available AI applications, already in use in radiology practices in European settings, was the most efficient way to introduce AI in their own clinical practices. This strategy was based on the belief that procuring off-the-shelf solutions would minimize the need to conduct comprehensive local validation processes in-house. In August 2021, a competitive dialogue procurement process commenced involving five qualified international vendors. At that time, each vendor highlighted that only a limited number of AI applications had been integrated into European clinical practices. These applications, tailored for specific use cases, demonstrated value within distinct clinical settings and for specific clinical tasks. This aligns with the study by Leeuwen and colleagues from 2021, which identified approximately 100 CE-marked algorithms that were regulatory approved and available on the market, with less than 10% having documented potential clinical effects. These solutions were approved primarily based on technical validation and diagnostic accuracy [2].

The insufficient documentation and limited evidence from deployments of the AI applications in real-world settings challenged the procuring hospital trust as they were to decide upon which AI applications to acquire. In response to this challenge, the hospital trust pivoted its initial plan from procuring individual AI applications from different vendors to buy a so-called marketplace or platform solution with several AI applications available. Procuring a platform solution was made possible as four of the five vendors presented this as an additional offer along with specific AI applications. By choosing the platform approach, the hospital trust saw an opportunity to be more flexible in terms of testing and potentially abandon the less documented applications in their local hospital practices while staying open for future AI applications.

3.1. Pre-implementation phase: Establishing trust among the radiologists.

This phase involved preparing for the final AI implementation. As the AI application about to be implemented was presented to the radiologists, it became evident that the lack of documentation also affected them. The radiologists expressed a certain skepticism towards the forthcoming deployment, especially related to whether the application was trained on representative data and whether it would apply to the hospitals' protocols for diagnostic practices.

Due to the uncertainty of the technological performance, the first AI application the hospital trust decided to implement was one of few technologies already in use in several European hospitals, namely BoneView from the vendor Gleamer. This AI application had extensive documentation supporting diagnostic accuracy. Since the area of analysis was bone fractures in X-rays, it was perceived as a tool with low risks of adverse outcomes. This choice coincided with the hospital trust's perception of establishing trust among clinicians as a crucial factor in succeeding with AI deployments in general. Thus, to build trust, the hospital trust carried out extensive prospective and retrospective validation of the application's diagnostic accuracy based on images from the local radiology practices before finally deciding to implement it in clinical practice.

From June to August 2023, the prospective validation of 1,600 skeletal images was analysed by the AI application for potential bone fractures. The results were compared with the results from traditional examinations of radiologists. In addition, the prospective validation acted as a proof-of-concept declaration regarding the technical set-up and the quality of the images. The results from the prospective validation were compared to the documentation of diagnostic accuracy provided by the vendor and previous research results. It became clear that the local validation did not detect differences in terms of sensitivity (number of false positives) and specificity (number of false negatives) compared with previous validation from other hospitals. Along with the confirmation of the existing documentation, the validation process also seemed to function as a way of gaining more trust among the clinicians as they got hands-on experience and/or insights into the application's performance and accuracy when analysing images from the hospital's patients.

3.2. Implementation phase: Organisational change management

The hospital trust's rationale for implementing AI applications was to meet the escalating demand for labour-intensive imaging examinations. Notably, the selected vendor emphasised that enhancing screening efficiency is not solely contingent on AI applications. Rather, the turning point revolves around organisational aspects and restructuring of current workflows and patient flows affected by the AI deployment. To facilitate the necessary changes, the vendor carried out a change management process involving several of the hospital trust's healthcare personnel. Additionally, the hospital trust employed two experienced radiologists with part-time roles as AI radiologists. These played a crucial role in supporting the entire change management and implementation process.

In short, the implementation at each hospital site started with workshops involving representatives from all actors affected by the upcoming AI deployment. At each of the four hospital locations, they thoroughly mapped the existing workflow and patient flows and identified the socio-technical changes necessary to address and mitigate in order to enhance the chances of success. The overarching changes involved alterations in the workflow of radiographers and radiologists. In the previous workflow, the radiographers received and prepared the patients and executed the X-rays to identify potential bone fractures. The radiographer also did the first assessment of the X-rays, but a radiologist was always responsible for the final patient diagnostics. However, the radiologists often needed to examine skeletal X-rays in between other examinations from other modalities. Consequently, patients frequently had to wait up to several hours before receiving the examination result and further instructions regarding treatment or discharge.

Within the new workflow, the radiographers, with the AI application as decision support, are given a greater responsibility in the initial diagnostic process to reduce patient waiting time. The alteration of responsibility implies that the radiographers use the results from the AI application to determine the patient's pathway. If the AI result shows "no fracture", the patient is discharged; if "fracture" is the outcome, the patient is transferred to an in-house unit for treatment and follow-up, while a result marked "doubtful" means that the patient must wait for a radiologist to examine the images. As a safety insurance, a radiologist always examines all the X-rays screened by AI as soon as possible.

Early evaluations of the AI implementation showed an improved patient pathway "because the patients don't have to wait so long". The reduced waiting time and discharge of patients with "no fracture" improved the workflow in the emergency room. However, the application did not directly reduce the radiologists' workload as they still had to examine all the X-rays.

4. Concluding Discussion

This section will present three challenges emerging during the procurement and implementation process and discuss how they were managed.

First, a key concern when deploying CE-marked AI solutions is that the performance of an application will vary across different contexts and geographies, clinical and patient sources [5]. The concern arises from a notable deficiency in the documentation accompanying the CE-marked algorithms, primarily comprising results derived from retrospective training and testing conducted within a development context [4]. In our study, the deficiency in documentation became a key challenge already during the procurement phase. This resulted in a shift in procurement strategy, where the easiest way forward was perceived as procuring a platform that potentially allowed choosing between a wider range of AI applications after settling the procurement. Hence, solving these challenges required a change in focus from procuring single market-ready AI solutions to acquiring and implementing an AI platform offering the hospital trust access to five AI applications [5].

Second, a significant challenge revolved around trust in the AI application caused by the deficiency in the documentation. The radiologists expressed skepticism regarding the algorithm's reliability in terms of diagnostic accuracy and its compatibility with the local information infrastructure (II). The lack of trust was met by conducting local validation of the algorithm. Establishing trust among clinicians constitutes a critical aspect of the social component of an II, in which trust represents a significant challenge to surmount when integrating technology as a new companion in the clinical decision-making process [7,8].

Third, the AI solution shaped the existing practices considerably. To overcome the challenges of re-organising clinical practices, an important preparatory measure for implementation entailed mapping existing workflows and patient flows as a baseline for the necessary changes. Re-organising and shaping practices also required communication and training of various groups of healthcare personnel. These changes are not necessarily a challenge per se, but these activities are vital for adapting new technology into an evolving II [8].

Deploying AI applications entails identifying the challenges and opportunities that arise throughout the process and addressing these issues as they emerge, which are not always foreseeable before initiation. To manage these emerging unforeseen challenges, we suggest that process evaluations of AI deployments in real-world settings play an essential role in ensuring a more sustainable outcome of digital transformations.

References

[1] Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. Npj Digit Med. 11. september 2020;3(1):1–8.

[2] van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol. juni 2021;31(6):3797–804.

[3] Your Europe [Internett]. [sitert 13. mars 2024]. CE marking – obtaining the certificate, EU requirements. Tilgjengelig på: https://europa.eu/youreurope/business/product-requirements/labels-markings/ce-marking/index_en.htm

[4] Oppenheimer J, Lüken S, Hamm B, Niehues SM. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. Life. januar 2023;13(1):223.

[5] Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, mfl. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. Lancet Digit Health. september 2021;3(9):e543–54.

[6] Ellingsen G, Silsand L, Severinsen GH, Linstad L. Scaling AI Projects for Radiology – Causes and Consequences. I: Challenges of Trustable AI and Added-Value on Health [Internett]. IOS Press; 2022 [sitert 8. mars 2024]. s. 13–7. Tilgjengelig på: https://ebooks.iospress.nl/doi/10.3233/SHTI220387

[7] Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, mfl. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res. 2017;19(11):e367.

[8] Hanseth O, Monteiro E. Understanding Information Infrastructure [Internett]. 1998 [sitert 4. oktober 2017]. Tilgjengelig på: http://heim.ifi.uio.no/~oleha/Publications/bok.pdf

[9] Klein HK, Myers MD. A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. 1. mars 1999;23(1):67–93.