Good Evaluation - Better Digital Health
U.H. Hübner et al. (Eds.)

doi:10.3233/SHTI251485

Evaluation of Clinical AI-Based Diagnostic Solutions – A Multiperspective, Interdisciplinary Approach

Thomas KAROPKA[a,1], Carina ØSTERVIG BYSKOV [b] and Martin DYRBA [c]
[a] *BioCon Valley GmbH*
[b] *Innovationsklinikken Aalborg Universitetshospital*
[c] *German Center for Neurodegenerative Diseases (DZNE)*
ORCiD ID: Thomas Karopka https://orcid.org/0000-0002-4448-6853,
Martin Dyrba https://orcid.org/0000-0002-3353-3167

Abstract. Introduction: The primary goal of developing new clinical diagnostic solutions is to create value for healthcare. The rapid rise of artificial intelligence (AI)-based diagnostics has led to a surge in publications and, to a lesser extent, market-ready tools. Clinicians must now integrate these innovations to manage increasing data volumes, making it challenging to assess the added value of new tools in the diagnostic workflow. Methods: The INTERREG Baltic Sea Region project "Clinical Artificial Intelligence-Based Diagnostics (CAIDX)" developed a comprehensive blueprint guiding the process from identifying clinical needs to implementing certified AI products in diagnostics. The approach emphasizes systematic evaluation at each development stage and throughout the AI solution's lifecycle, incorporating diverse stakeholder perspectives and a range of evaluation methodologies. Results: The CAIDX project produced the "Clinical AI-Pathway," an end-to-end framework for integrating AI-based diagnostic tools. This framework provides methodologies and tools for systematic evaluation at all stages, ensuring alignment with clinical needs and rigorous assessment of value. Conclusions: Systematic, multi-perspective evaluation is crucial for successfully integrating AI diagnostics into clinical practice. The "Clinical AI-Pathway" framework offers a structured method for assessing and implementing AI solutions, supporting their value-driven adoption in healthcare. The framework, available at ClinicalAI.eu, aims to facilitate broader and more effective use of AI in clinical diagnostics.

Keywords. Artificial Intelligence, evaluation, implementation, clinical diagnostics

1. Introduction

The integration of artificial intelligence (AI) into medical diagnostics represents a paradigm shift in healthcare, offering unprecedented opportunities to enhance diagnostic accuracy, streamline workflows, and personalize patient care. However, the rapid proliferation of AI-based diagnostic solutions necessitates rigorous evaluation frameworks to ensure safety, efficacy, effectiveness, and ethical alignment. Modern AI evaluation frameworks in healthcare emphasize holistic assessment beyond technical

[1] Corresponding Author: Thomas Karopka, BioCon Valley® GmbH, Markt 13, 17485 Greifswald.

performance. Critical challenges include balancing rigorous standards with flexibility, addressing multidisciplinary training gaps, and ensuring equity across diverse healthcare settings. Emerging theoretical frameworks prioritize scalability, transparency, and real-world interoperability while advocating for stakeholder collaboration across developers, clinicians, and policymakers [1].

The Clinical AI-based Diagnostics (CAIDX) [2] project has been conceived to provide a comprehensive framework and toolkit for the development and implementation of AI-based clinical solutions, available at ClinicalAI.eu. This framework spans the entire process, from initial clinical need identification to routine clinical application. The toolkit integrates both internally developed materials and external tools, with this publication focusing primarily on the latter. These tools are designed not only for evaluation purposes but also to facilitate systematic and structured development.

The CAIDX framework addresses a critical gap in the field, bridging the gap between AI solution providers and healthcare institutions. By providing standardized procedures for collaboration in development, testing, and implementation, coupled with guidance on regulatory and ethical considerations, CAIDX seeks to accelerate the adoption of AI technologies in clinical diagnostics. This paper provides a comprehensive overview of publicly available resources to help stakeholders navigate this challenge.

2. Methods

A comprehensive review of the existing literature and expert interviews were conducted to ascertain best practices and to identify tools that are already available. To achieve this, the project partners interviewed n=20 AI stakeholders from various regions (Denmark, Finland, Sweden, Germany, Poland, and Estonia) who were responsible for the development of products from the initial idea and prototype stage up to CE marking.

The interviewees constituted a diverse group of key stakeholders from the healthcare sector, including CEOs, CSOs, senior scientists, heads of clinical IT operations, and clinicians. Interviews were recorded and qualitatively analyzed. The results of the interviews will be published in a companion paper which is under development.

For the systematic literature review, the methodology employed was that of a snowball sampling approach, initiated with a Google, Google Scholar, and Web-of-Science search for reporting frameworks for AI studies, and augmented by the integration of the perspectives of change management, implementation science, and ethical frameworks within the context of AI in healthcare. All initial results were screened for public availability, the main topic of the tool/framework, and practical relevance, in order to create a directory of resources.

3. Results

For the organization of the identified frameworks and tools we assigned a category according to their main purpose. The categories are: Data, AI Security, Reporting, Ethics, Evaluation, Trustworthy AI, Implementation, Legal, Regulatory, and Health Technology Assessment (HTA) for AI. Table 1 summarizes the identified tools. We refer the reader to ClinicalAI.eu/read-also/ for an up-to-date version of this overview.

Table 1. Overview of tools and publications to evaluate and support implementation of AI-based solutions

Category	Tool	Author	Ref.
Data	SEMLA	Alexandersson et al. 2022	[3]
	Self-assessment guide for AI	CNIL 2022	[4]
	GDPR2DSM	TIEKE 2022	[5]
	METRIC framework	Schwabe et al. 2024	[6]
	Good machine learning practice	FDA 2023	[7]
AI Security	OWASP AI Exchange	OWASP.ORG 2025	[8]
	MITRE ATLAS	The MITRE Corp. 2025	[9]
Reporting	Equator Network	The EQUATOR Network 2025	[10]
Ethics	Ethics guidelines for trustworthy AI	AI High Level Expert Grp 2019	[11]
	Z-Inspection®	Zicari et al. 2021	[12]
Evaluation	MI-CLAIM checklist	Norgeot et al. 2020	[13]
	CLAIM checklist	Tejani et al. 2024	[14]
	ECLAIR guidelines	Omoumi et al. 2021	[15]
	Multi-Society statement radiology	Brady et al. 2024	[16]
	AI Literacy assessment matrix	Knoth et al. 2024	[17]
Trustworthy AI	Trustworthy AI development	de Manuel et al. 2024	[18]
	Trustworthy AI-enabled CDSS	Labkoff et al. 2024	[19]
	FUTURE-AI framework	Lekadir et al. 2025	[20]
Implementation	SALIENT Framework	van der Vegt et al. 2023	[21]
	Introductory Guide AI in Radiology	Warren et al. 2024	[22]
	Clinical Practice Integration of AI	Farrow et al. 2024	[23]
	From Bit to Bedside	Higgins & Madai 2020	[24]
	Dutch guideline for diagnostic AI in health	van Smeden et al. 2022	[25]
	AI for IMPACTS framework	Jacob et al. 2025	[26]
Legal	AI Act Risk Navigator	TÜV AI.LAB 2024	[27]
	AI Act Explorer	Future of Life Institute 2025	[28]
Regulatory	Questionnaire AI in medical devices	German Notified Bodies Alliance 2024	[29]
	Guideline for AI based Medical Devices	Johner Institute 2024	[30]
	Free EU MDR Templates	OpenReg GmbH 2024	[31]
HTA for AI	Health Technology Assessment for AI	Di Bidino et al. 2024	[32]
	MAS-AI	Fasterholdt et al. 2022	[33]
	HTA of AI-based medical devices	Farah et al. 2024	[34]

4. Discussion

The development of high-quality AI-based tools is contingent upon rigorous evaluation and constant monitoring. In this regard, the research community has developed a plethora of supportive tools, including reporting frameworks, checklists, and guidelines. These instruments address specific domains within the developmental process, encompassing aspects such as the reporting of clinical trial outcomes, e.g. TRIPOD-AI [35], DECIDE-AI [36] and CONSORT-AI [37], ethical evaluation [12] and the assessment of clinical applicability [13–16]. A number of tools have been developed to address every stage of the AI lifecycle, from identifying clinical needs to post-market surveillance [20,21,24–26]. HTA for AI has been addressed by Di Bidino et al. [32], Fasterholdt et al. [33], and Farah et al. [34].

In our collection of supportive tools, we have also included tools related to data access [3], data quality assessment [6] and AI Security [8,9]. Finally, tools for navigating the legal framework [27,28] and guidelines for developing AI based Medical Devices have been referenced [29–31]. In CAIDX, we are working on additional tools to facilitate matchmaking of project stakeholders and to enhance the understanding of tasks and roles in the development process. Results are available on the project website ClinicalAI.eu.

5. Conclusions

The proliferation of AI-based solutions for clinical diagnostics presents practitioners with the challenge of selecting AI tools that create genuine value and are predicated on a medical necessity. The ideal scenario would involve the co-development of these tools as a collaborative effort between medical and data science professionals. In the context of the CAIDX project, we advocate an even broader approach by recommending the involvement of all stakeholders involved in the value-creation pipeline, including business, IT and procurement specialists, with the objective of optimizing the probability of successful product availability to a large patient population.

It is imperative to emphasize the necessity of evaluating the solution in each stage of development and implementation from multiple perspectives. It is also emphasized that the development process is not linear but iterative, with cycles. The provision of metrics and tools for each step is beyond the scope of the CAIDX project and needs to be addressed in future. Consequently, we have collated a range of freely available tools for the identified perspectives that can be utilized along the lifecycle of an AI-based solution from development and implementation to post-market surveillance.

References

[1] Cresswell K, Keizer N de, Magrabi F, Williams R, Rigby M, Prgomet M, et al. Evaluating Artificial Intelligence in Clinical Settings—Let Us Not Reinvent the Wheel J Med Internet Res. 2024 Aug 7;26(1):e46407.
[2] CAIDX [Internet] Interreg Baltic Sea Region. [cited 2025 Jan 28]. Available from: https://interreg-baltic.eu/project/caidx/
[3] Alexandersson J, Britz J, Seimetz V, Tabellion D. An On-Premises Trusted Research Environment for AI-based R&D with Sensitive Personal Information. Available from: https://semla.dfki.de/white-paper/
[4] CNIL. Self-assessment guide for artificial intelligence (AI) systems [Internet] [cited 2024 Jun 13]. Available from: https://www.cnil.fr/en/self-assessment-guide-artificial-intelligence-ai-systems
[5] TIEKE. The Data Protection Tool GDPR2DSM [Internet] [cited 2024 Jul 4]. Available from: https://www.tietosuojaapkyrityksille.fi/en/
[6] Schwabe D, Becker K, Seyferth M, Klaß A, Schaeffter T. The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review Npj Digit Med. 2024 Aug 3;7(1):1–30.
[7] Health C for D and R. Good Machine Learning Practice for Medical Device Development: Guiding Principles FDA [Internet]. 2023 Oct 20 [cited 2024 Jan 15]; Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
[8] OWASPAI.ORG. AI Exchange [Internet] [cited 2025 Feb 4]. Available from: https://owaspai.org/
[9] The MITRE Corporation. MITRE ATLAS™ [Internet] [cited 2025 Feb 4]. Available from: https://atlas.mitre.org/
[10] The EQUATOR Network. EQUATOR Network | Enhancing the QUAlity and Transparency Of Health Research [Internet] [cited 2025 Feb 14]. Available from: https://www.equator-network.org/
[11] Directorate-General for Communications Networks, Content and Technology (European Commission). Ethics guidelines for trustworthy AI [Internet] Publications Office of the European Union; 2019 [cited 2025 Feb 14]. Available from: https://data.europa.eu/doi/10.2759/346720
[12] Zicari RV, Brodersen J, Brusseau J, Düdder B, Eichhorn T, Ivanov T, et al. Z-Inspection®: A Process to Assess Trustworthy AI IEEE Trans Technol Soc. 2021 Jun;2(2):83–97.
[13] Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist Nat Med. 2020 Sep;26(9):1320–4.
[14] Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update Radiol Artif Intell. 2024 Jul;6(4):e240300.
[15] Omoumi P, Ducarouge A, Tournier A, Harvey H, Kahn CE, Louvet-de Verchère F, et al. To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines) Eur Radiol. 2021 Jun;31(6):3786–96.

[16] Brady AP, Allen B, Chong J, Kotter E, Kottler N, Mongan J, et al. Developing, Purchasing, Implementing and Monitoring AI Tools in Radiology: Practical Considerations. A Multi-Society Statement from the ACR,CAR, ESR, RANZCR and RSNA Radiol Artif Intell. 2024 Jan;6(1):e230513.
[17] Knoth N, Decker M, Laupichler MC, Pinski M, Buchholtz N, Bata K, et al. Developing a holistic AI literacy assessment matrix – Bridging generic, domain-specific, and ethical competencies Comput Educ Open. 2024 Jun 1;6:100177.
[18] de-Manuel-y-Vicente C, Fernández-Narro D, Blanes-Selva V, García-Gómez JM, Sáez C. A Development Framework for Trustworthy Artificial Intelligence in Health with Example Code Pipelines [Internet] medRxiv; 2024 [cited 2024 Jul 19]. p. 2024.07.17.24310418. Available from: https://www.medrxiv.org/content/10.1101/2024.07.17.24310418v1
[19] Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support J Am Med Inform Assoc. 2024 Sep 26;ocae209.
[20] Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare BMJ. 2025 Feb 5;388:e081554.
[21] van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework J Am Med Inform Assoc JAMIA. 2023 May 19;ocad088.
[22] Warren BE, Bilbily A, Gichoya JW, Chartier LB, Fawzy A, Barragán C, et al. An Introductory Guide to Artificial Intelligence in Interventional Radiology: Part 2: Implementation Considerations and Harms Can Assoc Radiol J. 2024 Aug 1;75(3):568–74.
[23] Farrow L, Meek D, Leontidis G, Campbell M, Harrison E, Anderson L. The Clinical Practice Integration of Artificial Intelligence (CPI-AI) framework: a proposed application of IDEAL principles to artificial intelligence applications in trauma and orthopaedics Bone Jt Res. 2024 Sep 18;13(9):507–12.
[24] Higgins D, Madai VI. From Bit to Bedside: A Practical Framework for Artificial Intelligence Product Development in Healthcare Adv Intell Syst. 2020;2(10):2000052.
[25] Smeden M van, Moons KG, Hooft L, Chavannes NH, Os HJ van, Kant I. Guideline for high-quality diagnostic and prognostic applications of AI in healthcare 2022 Nov 14 [cited 2024 Sep 18]; Available from: https://osf.io/tnrjz/
[26] Jacob C, Brasier N, Laurenzi E, Heuss S, Mougiakakou SG, Cöltekin A, et al. AI for IMPACTS Framework for Evaluating the Long-Term Real-World Impacts of AI-Powered Clinician Tools: Systematic Review and Narrative Synthesis J Med Internet Res. 2025 Feb 5;27:e67485.
[27] TÜV AI.Lab GmbH. AI Act Risk Navigator [Internet] [cited 2024 Dec 10]. Available from: https://www.tuev-risk-navigator.ai
[28] Future of Life Institute. The AI Act Explorer | EU Artificial Intelligence Act [Internet] [cited 2024 Jun 13]. Available from: https://artificialintelligenceact.eu/ai-act-explorer/
[29] Sudmann V. Questionnaire Artificial Intelligence in medical devices. https://www.team-nb.org/wp-content/uploads/2024/11/Team-NB-PositionPaper-AI-in-MD-Questionnaire-V1-20241125.pdf
[30] Johner C. Guideline for AI based Medical Devices [Internet] 2024 [cited 2024 Dec 3]. Available from: https://github.com/johner-institut/ai-guideline
[31] OpenReg GmbH. Free EU MDR Templates [Internet] OpenRegulatory. [cited 2025 Jan 13]. Available from: https://openregulatory.com/templates/
[32] Di Bidino R, Daugbjerg S, Papavero SC, Haraldsen IH, Cicchetti A, Sacchini D. Health technology assessment framework for artificial intelligence-based technologies Int J Technol Assess Health Care. 2024 Nov 21;40(1):e61.
[33] Fasterholdt I, Kjølhede T, Naghavi-Behzad M, Schmidt T, Rautalammi QTS, Hildebrandt MG, et al. Model for ASsessing the value of Artificial Intelligence in medical imaging (MAS-AI) Int J Technol Assess Health Care. 2022 Jan;38(1):e74.
[34] Farah L, Borget I, Martelli N, Vallee A. Suitability of the Current Health Technology Assessment of Innovative Artificial Intelligence-Based Medical Devices: Scoping Literature Review J Med Internet Res. 2024 May 13;26(1):e51514.
[35] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods BMJ. 2024 Apr 16;e078378.
[36] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI BMJ. 2022 May 18;377:e070904.
[37] Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension The BMJ. 2020 Sep 9;370:m3164.