Practical considerations for developing AI use in mammography

Jarmo Reponen, MD, PhD, Professor, Radiologist

Research Unit of Health Sciences and Technology, Faculty of Medicine, Univ. of Oulu, Finland

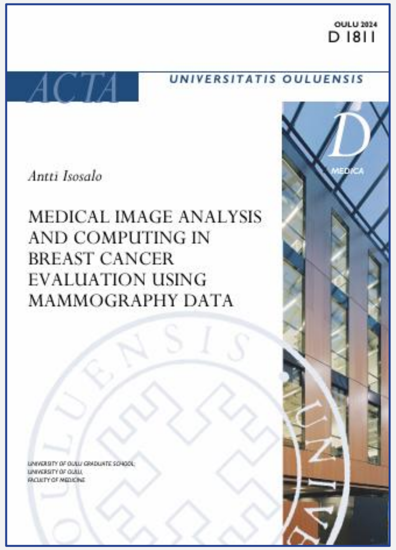
Antti Isosalo, PhD, MSc (Engineering),

Center for Wireless Communications, Faculty of Information Technology and Electrical Engineering, Univ. of Oulu, Finland

JSRS 2026, Copenhagen, Denmark 03.06.2026

Contents of the presentation

- Mammography screening practice
- Motivation for introducing AI
- Medical image analysis approaches in mammography
- Image phenotypes in human and computer-based interpretation
- Common challenges in AI interpretation
- Sources of bias
- Ideas and ideal AI workflow
- Practical research needs
- Maturity of AI in screening use



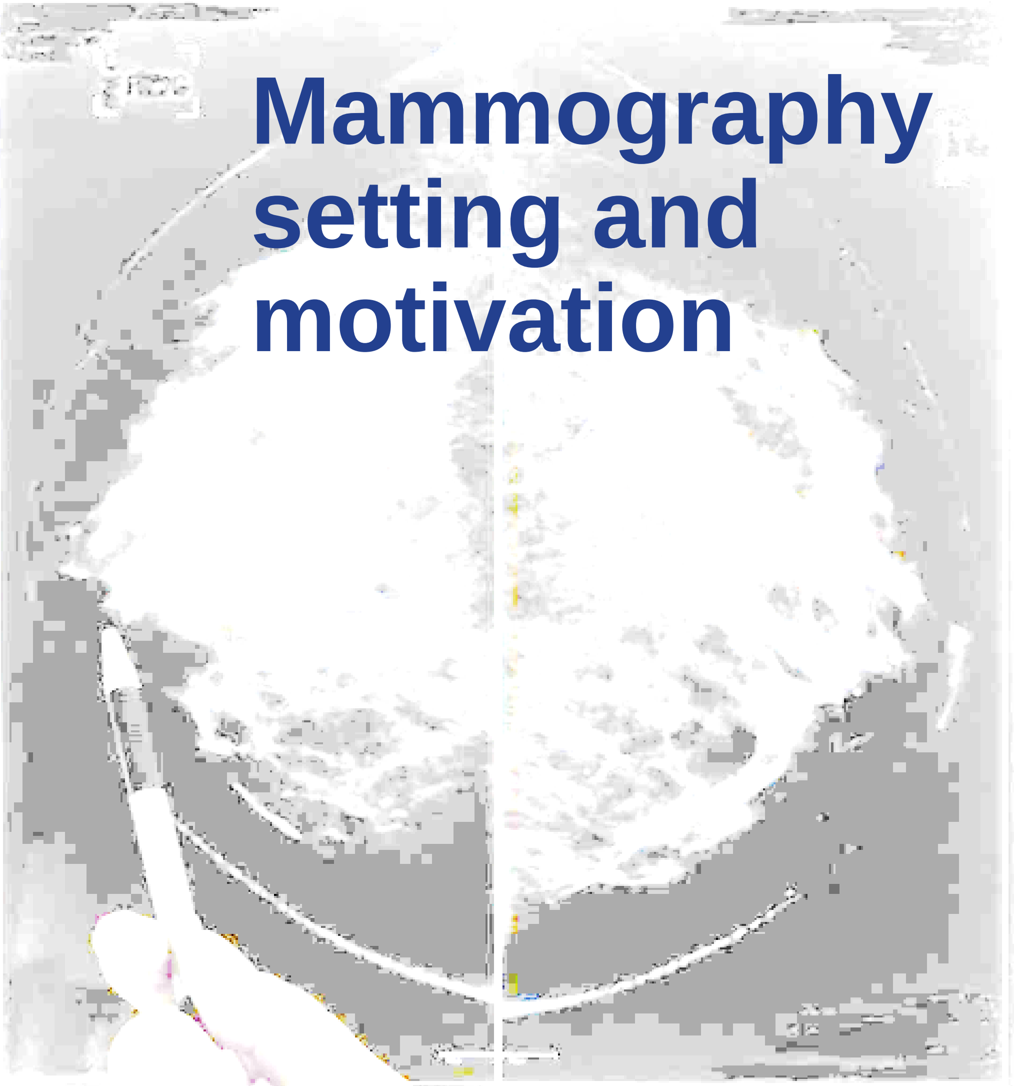
1. Isosalo A., Inkinen S. I., Turunen T., Ipatti P. S., Reponen J., & Nieminen M. T. (2023). Independent evaluation of a multi-view multi-task convolutional neural network breast cancer classification model using Finnish mammography screening data. Computers in Biology and Medicine 161, 107023.
2. Isosalo A., Mustonen H., Turunen T., Ipatti P. S., Reponen J., Nieminen M. T., & Inkinen S. I.(2022). Evaluation of different convolutional neural network encoder-decoder architectures for breast mass segmentation. In: Medical Imaging 2022: Imaging Informatics for Healthcare, Research, and Applications, 12037, International Society for Optics and Photonics, SPIE, 207–214.
3. Isosalo A., Inkinen S. I., Prostredná L., Heino H., Ipatti P. S., Reponen J., & Nieminen M. T. (2024). Imaging phenotype evaluation from digital breast tomosynthesis data: A preliminary study. Computers in Biology and Medicine 183, 109285.
4. Isosalo, A., Islam, J., Mustonen, H., Räinä, E., Inkinen, S. I., Brix, M., Kumar, T., Reponen, J., Nieminen, M. T., & Harjula, E. (2023). Local edge computing for radiological image reconstruction and computer-assisted detection: A feasibility study. Finnish Journal of eHealth and eWelfare, 15(1), 52–66.
5. 5. Isosalo A., Inkinen S. I., Heino H., Turunen T. & Nieminen M. T. (2024). Mammogram AnnotationTool: Markup tool for breast tissue abnormality annotation. Software Impacts 19, 100599.

https://urn.fi/URN:NBN:fi:oulu-202410106255

University of Oulu. (2021). Oulu dataset of screening mammography.
URL: http://urn.fi/URN:NBN:fi:att:30ebfbd2-8f95-4945-bc28-066b6f4a9cbd

-49,500 curated mammography screening examinations, 209,000 images with full MIS data. All 3000 further examined not normal cases were relabeled for AI training by radiologists.

Mammography setting and motivation

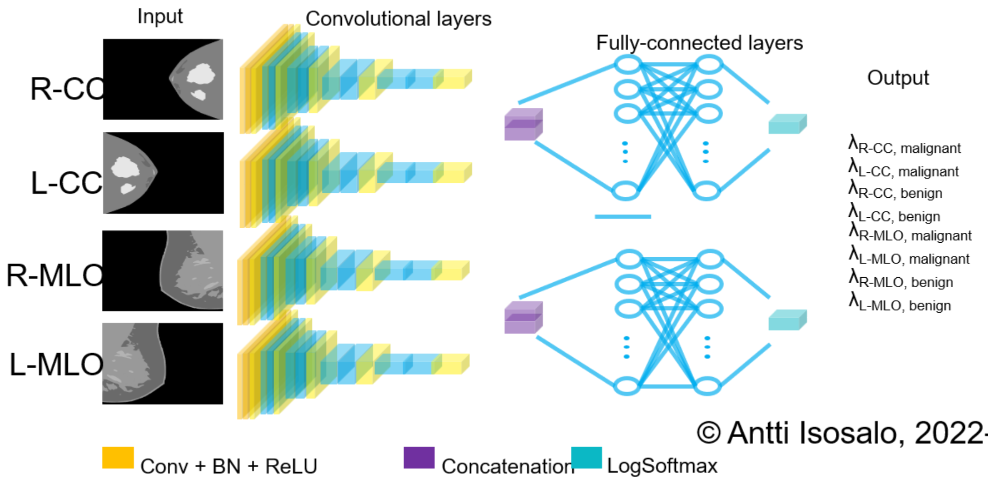


- **Screening Scale and Challenges:**
  Population based activity, mammography screening involves millions of exams yearly with low cancer prevalence and high normal image volume.
- **Radiologist Workload and Shortage:**
  Increased workload from aging populations and new tech like digital breast tomosynthesis, but limited specialized radiologist resources.
- **Motivation: AI as Support Tool?**
  AI could support screening by reducing workload (Triage), enhancing consistency, and aiding double-reading strategies.
- **Critical Screening Requirements:**
  Screening demands high sensitivity, safety, and transparency due to population-level impact of errors. AI is not a replacement for humans but a response to systemic challenges.

Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al.. (2019) Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol. 2019 Sep;29(9):4825-4832. doi: 10.1007/s00330-019-06186-9

Nishikawa, R. M., & Lu, A. H. (2023). AI in Screening Mammography: Use One Radiologist and Improve Double Reads. Radiology, 309(2), e232964. https://doi.org/10.1148/radiol.232964

Medical image analysis approaches in mammography

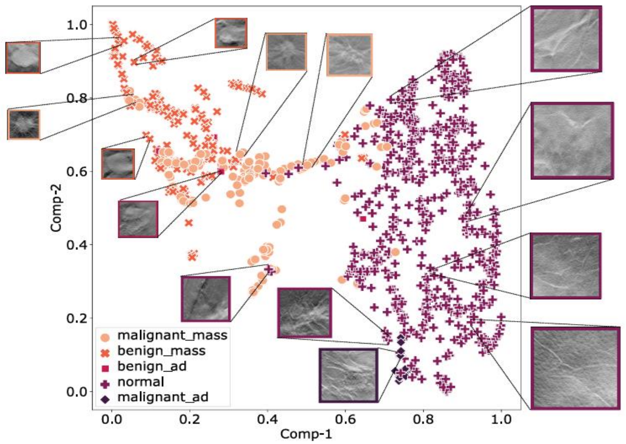


Amari, S. (1967). A theory of adaptive pattern classifiers. IEEE Transactions on Electronic Computers 16(3), 299-307. https://doi.org/10.1109/PGEC.1967.264666

Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. IEEE Transactions on Systems, Man, and Cybernetics 13(5), 826-834. https://doi.org/10.1109/TSMC.1983.6313076

- Early/ Segmentation and Features: Initial mammography analysis used handcrafted segmentation and feature extraction describing shape, intensity, and texture.
- Radiomics for Tissue Quantification: Radiomics quantifies tissue characteristics in a reproducible, high-dimensional manner to enhance lesion description.
- Now/ Deep Learning and CNNs: Deep learning uses convolutional neural networks to learn hierarchical representations directly from mammography data.
- Transfer Learning: allows models pretrained on large datasets to be adapted to mammography tasks
- Hybrid Models: Combining deep learning with radiomic or clinical features

Image phenotypes in human and computer-based interpretation



© Antti Isosalo, 2024

** Tabar L, Dean P, et al. (2023). Imaging biomarkers are underutilised but highly predictive prognostic factors for the more fatal breast cancer subtypes. European Journal of Radiology 166, 111021. https://doi.org/10.1016/j.ejrad.2023.111021

– Human Imaging Phenotypes:
  Radiologists interpret mammograms by identifying lesion density, shape, margins, and background tissue patterns guiding diagnosis. Re: L. Tabar **.

– Computational Radiomic Phenotypes:
  Computers quantify texture, intensity, spatial relationships, and shape enabling machine learning and statistical modeling. > The underlying goal remains the same: to characterize tissue in a way that correlates with pathology

– Human vs Computer Interpretation:
  Human phenotypes are holistic and context-dependent, while computational phenotypes require precise, consistent data definitions.

– Bridging the Interpretation Gap:
  Aligning AI systems with clinical reasoning is essential to build trust and improve diagnostic accuracy.

Common challenges in AI interpretation

- **Class Imbalance Impact:** Cancer cases are rare in screening data, causing AI models to miss critical clinical findings despite high overall accuracy.
- **Hidden Stratification Issues:** Rare but important subgroups like specific breast densities or lesion types may be underrepresented, complicating AI evaluation. (Compare human with rare cases!)
- **Generalization Limitations:** AI models often fail to generalize well across different institutions, vendors, and populations.
- **Interpretability Concerns:** Deep learning models produce accurate predictions but lack transparent explanations for clinical assessment.
- **Poor reference standards:** e.g.interval cancers, imperfect ground truth labels affect training and evaluation.

Al-Karawi, D., Al-Zaidi, S., Helael, K. A., Obeidat, N., Mouhsen, A. M., Ajam, T., Alshalabi, B. A., Salman, M., & Ahmed, M. H. (2024). A Review of Artificial Intelligence in Breast Imaging. Tomography (Ann Arbor, Mich.), 10(5), 705–726. https://doi.org/10.3390/tomography10050055

Sources of bias in AI interpretation

- **Data Collection Bias:** Historical practices and selective data inclusion create bias in AI training data reflecting real-world imbalances.
- **Population Sampling Bias:** Training datasets may underrepresent key demographics like age groups, breast densities, and ethnic backgrounds.
- **Device and Protocol Variability:** Differences in imaging devices and protocols cause systematic variations that models may learn incorrectly.
- **Annotation and Automation Bias:** Inter-reader variability and overreliance on AI outputs introduce subjective and automation biases in interpretation.
- **Imaging procedures caused bias:** Signs of medical intervention in the data, e.g., a breast tissue marker appearing in a mammogram, can lead to issues in AI training.

Drukker et al. (2023). Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. Journal of medical imaging 10(6), 061104. https://doi.org/10.1117/1.JMI.10.6.061104

Varoquaux, G., & Cheplygina, V. (2022, April). Machine learning for medical imaging: Methodological failures and recommendations for the future. Digital Medicine 5(48), 1-8. https://doi.org/10.1038/s41746-022-00592-y

Development ideas and ideal AI workflow

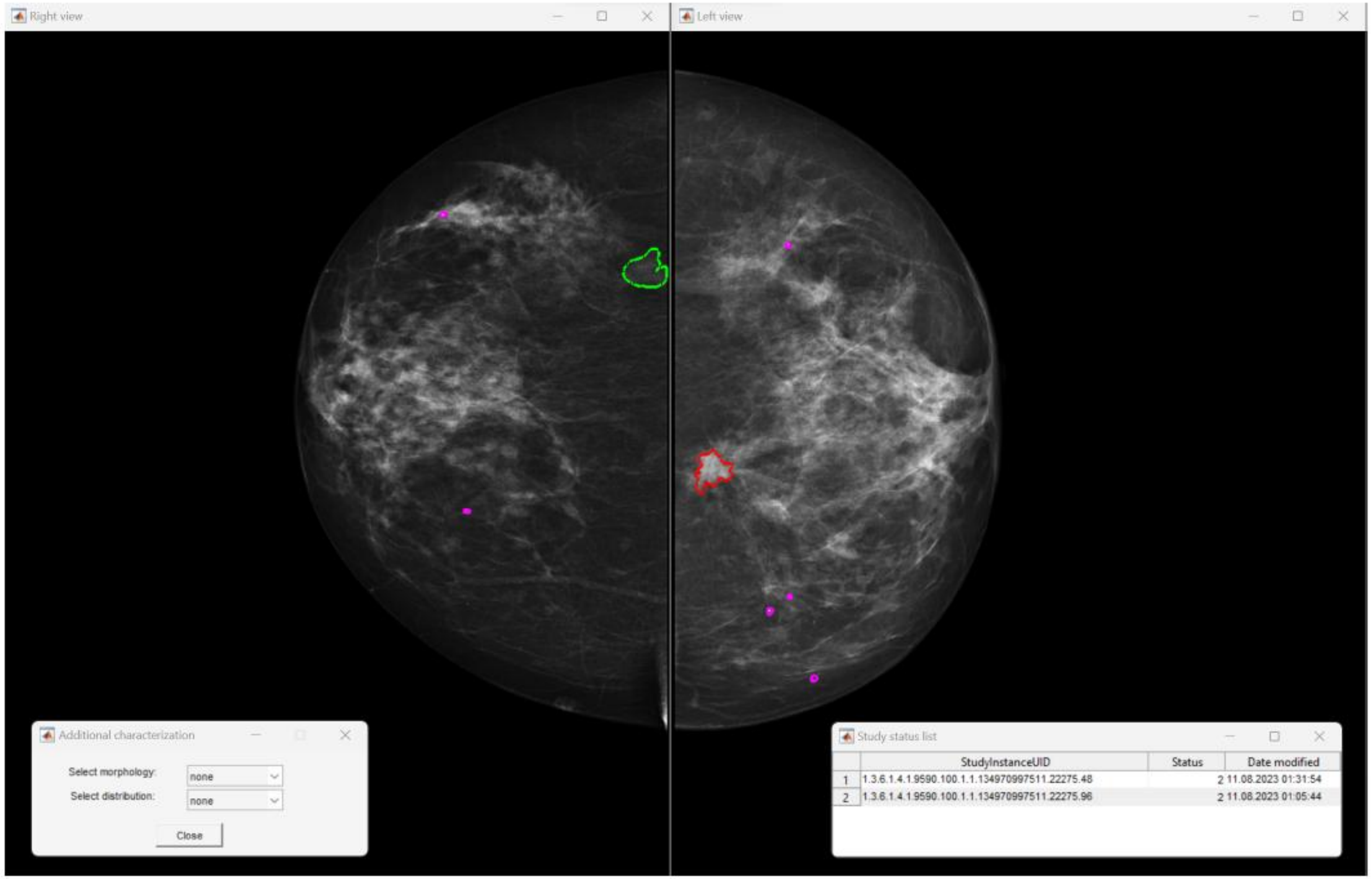
- **Dataset Curation and Annotation:** Ensuring representativeness and quality in datasets via metadata completeness and expert-supported annotation improves AI reliability.
- **Validation and Evaluation:** External and multi-center validation assess AI generalization and robustness across diverse clinical settings.
- **Human-AI Collaborative Workflow:** AI serves as a supportive tool like triage or second reader, maintaining radiologist responsibility and situational awareness. Ai is not a stand-alone decision maker.
- **Continuous Monitoring and Feedback:** Ongoing monitoring and feedback enable AI workflow evolution and integration into clinical screening practice.

Drukker et al. (2023). Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. Journal of medical imaging 10(6), 061104. https://doi.org/10.1117/1.JMI.10.6.061104

Example: Mitigating the annotation bias

Annotation bias can be alleviated by standardizing the process, e.g., by using a user interface tailored for the task.

Isosalo A., Inkinen S. I., Heino H., Turunen T. & Nieminen M. T. (2024). Mammogram AnnotationTool:
Markup tool for breast tissue abnormality annotation. Software Impacts 19, 100599.
https://doi.org/10.1016/j.simpa.2023.100599.



Practical research needs when introducing AI to mammography

Yalda Zafari, Hongyi Pan, Gorkem Durak, Ulas Bagci, Essam A. Rashed, and Mohamed Mabrok (2025) MammoClean: Toward Reproducible and Bias-Aware AI in Mammography through Dataset Harmonization.https://arxiv.org/html/2511.02400v1

Etsuji Nakai, Yumi Miyagi, Kazuhiro Suzuki, Alessandro Scoccia Pappagallo, Hiroki Kayama, Takehito Matsuba, Lin Yang, Shawn Xu, Christopher Kelly, Ryan Najafi, Timo Kohlberger, Daniel Golden, Akib Uddin, Yusuke Nakamura, Yumi Kokubu, Yoko Takahashi, Takayuki Ueno, Masahiko Oguchi, Shinji Ohno, Joseph R Ledsam, (2024) Artificial intelligence as a second reader for screening mammography, Radiology Advances, Volume 1, Issue 2, umae011, https://doi.org/10.1093/radadv/umae011

- Robust, Diverse Data for AI Training:
  Robust AI training requires large, well-curated, and diverse datasets to ensure reliable mammography screening.
- Explainable AI Methods:
  Improving explainability and interpretability of AI aligns outputs with clinical reasoning and builds trust.
- Integration with Clinical Workflow:
  Research must evaluate how AI interacts with radiologists, workflow efficiency, and organizational structures.
- Ethical and Regulatory Focus:
  Attention to regulatory, ethical, and medico-legal issues is crucial for safe AI adoption in population screening.

Scandinavian landmark study of AI in screening practice

Lång et al. (2023) **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): 936-944. https://doi.org/10.1016/S1470-2045(23)00298-X

- The MASAI study, world's first RCT to evaluate AI in mammography screening. It compared AI-assisted screening reading with traditional double reading by two radiologists in more than 80,000 Swedish women participating
- A digital mammography system from a single manufacturer, and AI analysis using the Transpara 1.7.0 system, which has been trained on over 200,000 mammographic examinations from multiple countries and different device environments.
- The AI-assisted workflow proved to be clinically safe and detected approx. 20% more cancers without increasing the number of false positives. The workload of screening interpretation was reduced, by about 44%, as most examinations could be assigned to a single radiologist.
- The MASAI study demonstrates: the clinical benefit of AI in mammography screening emerges only when these methodologically and clinically distinct tasks—such as abnormality detection, classification, and prioritization of reading—are systematically integrated into the screening process.

Conclusions: Maturity aspects of AI in screening mammography

- Technical Maturity and Benefits:
  AI in mammography can offer workload reduction and decision support, showing meaningful clinical benefits.
- Current Limitations:
  Bias, generalization challenges, interpretability and workflow integration issues limit AI adoption in screening.
- Collaborative Reading Strategy:
  AI performs best when complementing radiologists, supporting a collaborative screening approach.
- Responsible Deployment:
  Successful AI deployment requires validation, transparency, monitoring, and alignment with clinical values.

Further Reading

- Lauritzen, A. D., Lillholm, M., Lynge, E., Nielsen, M., Karssemeijer, N., & Vejborg, I. (2024). Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. *Radiology*, *311*(3), e232479. https://doi.org/10.1148/radiol.232479
- Elías-Cabot, E., Romero-Martín, S., Raya-Povedano, J. L., Brehl, A. K., & Álvarez-Benito, M. (2024). Impact of real-life use of artificial intelligence as support for human reading in a population-based breast cancer screening program with mammography and tomosynthesis. European radiology, 34(6), 3958–3966. https://doi.org/10.1007/s00330-023-10426-4
- van Winkel, S. L., Peters, J., Janssen, N., Kroes, J., Loehrer, E. A., Gommers, J., Sechopoulos, I., de Munck, L., Teuwen, J., Broeders, M., Karssemeijer, N., & Mann, R. M. (2025). AI as an independent second reader in detection of clinically relevant breast cancers within a population-based screening programme in the Netherlands: a retrospective cohort study. The Lancet. Digital health, 7(8), 100882. https://doi.org/10.1016/j.landig.2025.10088
- Marinovich, M. L., Lotter, W., Waddell, A., & Houssami, N. (2025). Simulated arbitration of discordance between radiologists and artificial intelligence interpretation of breast cancer screening mammograms. Journal of medical screening, 32(1), 48–52. https://doi.org/10.1177/09691413241262960
- Lehnen, T., Polenske, D., Wichtmann, B. D., & Lehnen, N. C. (2026). AI software as a third reader in breast cancer screening-a prospective diagnostic observational study. European radiology, 36(6), 4478–4488. https://doi.org/10.1007/s00330-026-12359-0

Arigato! Takk! Kiitos! Thank you!

Contact: Jarmo Reponen
e-mail: jarmo.reponen@oulu.fi

Thanks to the Oulu Medical Data Infrastructure (OMDI) project working group:

Outi Laatikainen (Clinical Pharmacologist) Antti Isosalo (Software Engineering), Pekka Siirtola (Information Processing), Satu Tamminen (Information Processing), Anusha Ihalapathirana (Information Processing), Gunjan Chandra (Information Processing), Miika Nieminen (Medical Physics), Santeri Rytky (Medical Physics), Joonas Kujala (Medical Physics) Pieta Ipatti (Radiologist), Topi Turunen (Radiologist), Taneli Mattila (Pathologist), Hanne Kuitunen (Oncologist), Jarmo Reponen (Radiologist).

OULUN
YLIOPISTO