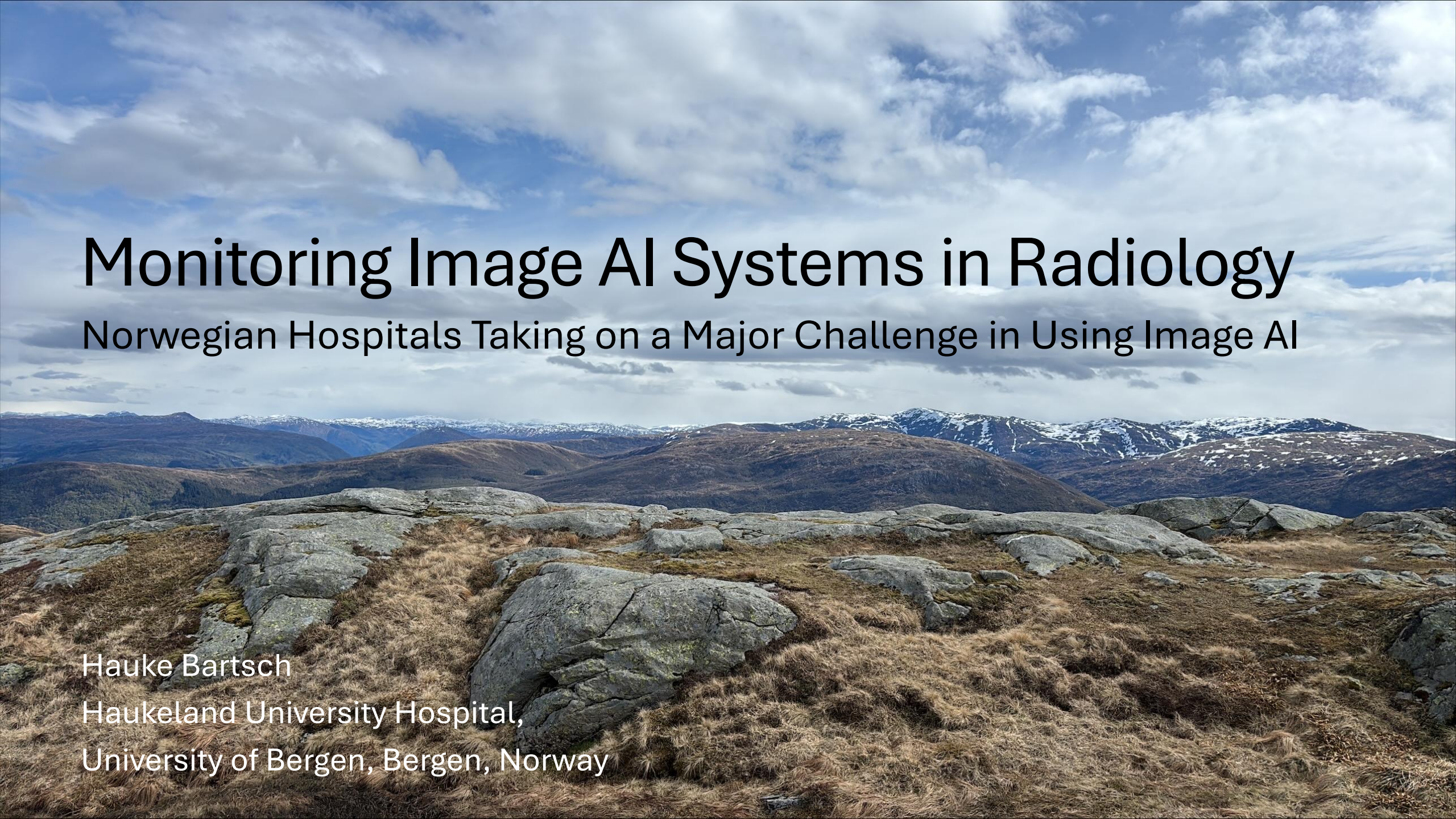
Monitoring Image AI Systems in Radiology

Norwegian Hospitals Taking on a Major Challenge in Using Image AI



Hauke Bartsch
Haukeland University Hospital,
University of Bergen, Bergen, Norway

AI as Clinical Infrastructure in Radiology

AI Integration in Radiology

AI systems support radiologists by detecting abnormalities and by providing better measurements.

AI as Clinical Infrastructure

AI functions as essential clinical infrastructure, they enhance existing data generation and data processing systems, critical for patient outcomes.

Continuous Monitoring Importance

Ongoing monitoring is vital to ensure AI reliability, adapting to data changes and software updates over time.

Responsibility in Public Healthcare

Norwegian public hospitals emphasize quality, equity, and safety in AI deployment and ongoing evaluation.

Commercial AI system in Breast Screen Norway (PaaS)
Transpara aids radiologists by highlighting suspicious mammogram regions and providing risk scores for breast cancer screening

In-house Build AI for Detection of Endometrial Cancer
Clinical integration of machine learning derived tumor size and radiomic tumor profiles in cancer diagnostics
E. Hodneland

Why AI Systems Change Over Time

Technical, Data, and Workflow-Related Changes

Impact of Software Updates

Regular AI model updates can alter performance by changing training data, model architecture, or decision thresholds.

Data Drift Effects

Changes in imaging equipment or patient demographics can cause shifts in input data, reducing AI accuracy over time.

Workflow Integration Changes

Adjustments in how clinicians use AI outputs or system integration impact clinical usage and diagnostic outcomes.

Technical and Hardware Issues

DICOM inconsistencies, preprocessing errors, and hardware changes can subtly affect AI diagnostic behavior.

Why Monitoring Is Necessary

Patient Safety, Quality Assurance, and Trust

Importance of Monitoring AI

Consistent monitoring ensures AI systems maintain accuracy.

Impact of AI Performance Changes

Small changes in sensitivity or specificity can lead to missed diagnoses or increased recall rates, affecting patient outcomes.

Building Trust and Compliance

Reliable AI behavior fosters clinician trust and supports regulatory compliance and ethical standards in healthcare.

Examples of AI Systems in Clinical Use

1

Commercial AI system in Breast Screen Norway (PaaS)

Transpara aids radiologists by highlighting suspicious mammogram regions and providing risk scores for breast cancer screening.

- Monitoring and Stability
  - Monitoring ensures consistent recall rates, stable score distributions, and comparable performance across hospitals and time

- Quality Control and Validation
  - CE-marked AI systems require local monitoring and silent mode testing after upgrades to maintain quality control.

2

In-house developed AI system for clinical use (Research)

Clinical integration of machine learning derived tumor size and radiomic tumor profiles in cancer diagnostics.

- Integration into clinical PACS (Helse Vest RHF)
  - PACS initiated analysis
  - Read of AI generated structured reports in PACS
- Establish consistent reporting for AI results in PACS
  - DICOM aware governor module (contrast, age, sex, MR sequence)
  - Interpretable uncertainty using vote maps
  - Versioning of AI and integration system

Commercial AI: Transpara for Breast Cancer Screening

Challenges: Commercial AI systems are frequently updated

- Monitoring and Stability
    - Monitoring ensures consistent recall rates, stable score distributions, and comparable performance across hospitals and time.
- Quality Control and Validation
    - Even CE-marked AI systems require local monitoring and silent mode testing after upgrades to maintain quality control.

In-House AI: Endometrial Cancer Detection in MRI

Challenges of In-House AI

Smaller datasets and variability in imaging protocols. But training data is fully known (possibility for digital twin setup)

Responsibility for Validation

Hospitals bear full responsibility for validating and maintaining in-house AI systems ensuring clinical alignment and safety.

Continuous Performance Evaluation

Regular audits and outcome comparisons improve AI accuracy by analyzing false positives and negatives continuously.

Governance and Long-Term Commitment

Strong governance structures and sustained monitoring ensure local AI innovation meets clinical standards and safety.

Norwegian Hospitals' Monitoring Approach

AI Monitoring and Quality Assurance Project (VaMos)

AI performance is monitored regularly using a vendor neutral validation system currently developed at Haukeland University Hospital.

Primary aim: compare AI malignancy score, diagnostic performance and test failure using the same sample

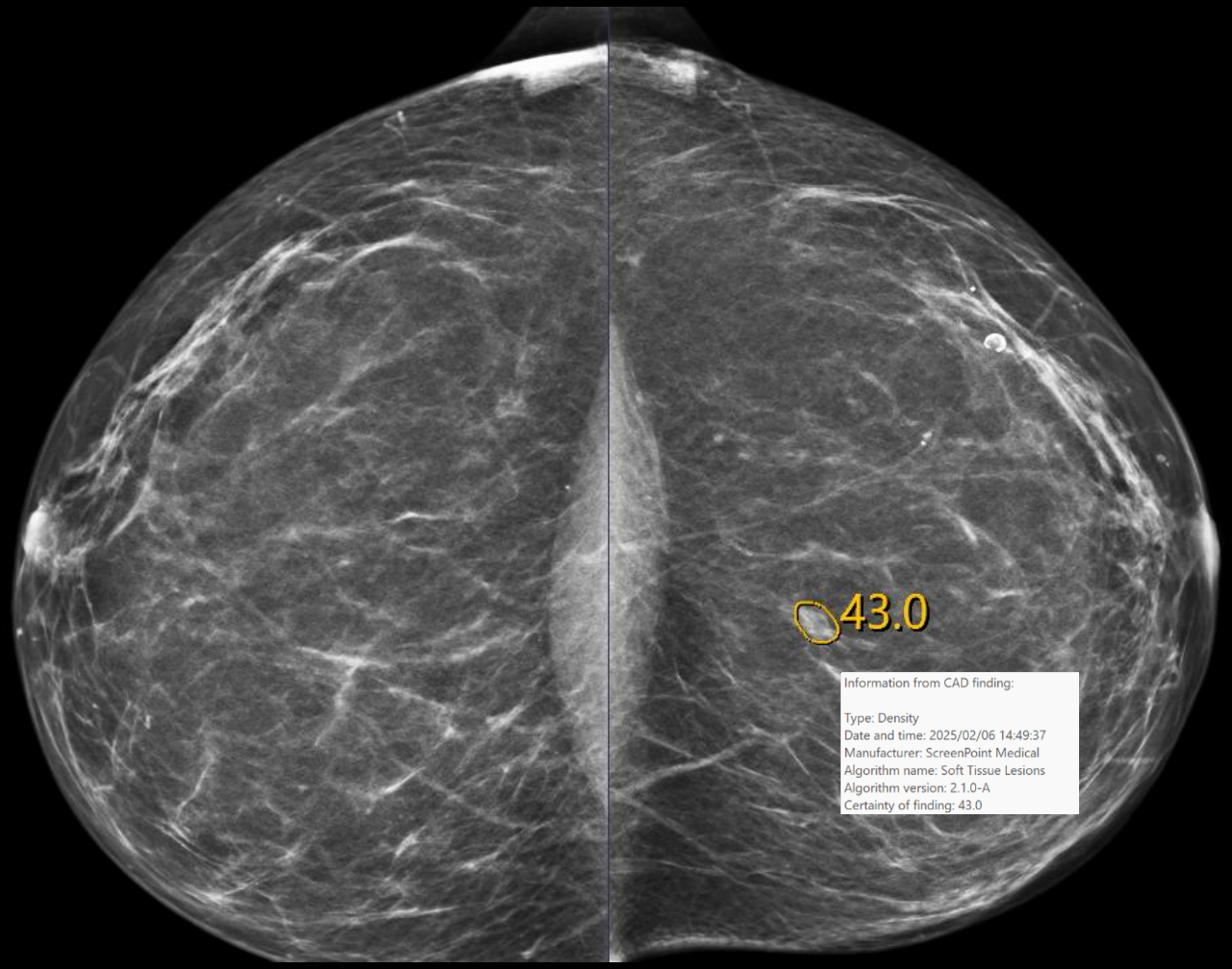
Secondary aim: explore impact of independent variables on AI performance like age, screening history, mammographic features and technical aspects.

Measures of interest:

- Cancer detection rates,
- recall rates,
- distribution of AI scores for suspicion of cancer and
- technical error rates

stratified by ethnicity, breast density, histopathology and mammography vendor / scan parameter.

Solveig Hofvind, Kristina Kymre; Marthe Larsen; Åsne Sørlien Holen, Kerstin Obst-Gleditsch, Hauke Bartsch, Leif Oltedal, Ingfrid Haldorsen



Workflow (consensus reads)

Reader 1 + Reader 2

Recall screening participant on findings

New workflow to reduce participants waiting times

Reader 1 + AI

Recall screening participant on findings

How can we trust the AI?

(monitor model performance and model drift)

VaMos will test if AI produces the same findings on the same data (N=1000), across different versions of the AI software run at different times (weekly).

VaMos will test how stable the recall rates are over time using random sample from current AIMS (N=6x1000, over 6 month).

Status: Ethical approval done, DPIA and ROS done, software development ongoing

Take-Home Messages

Continuous AI Monitoring

Ongoing monitoring of AI systems post-deployment maintains patient safety and clinical performance stability.

Dynamic AI Behavior

AI can change due to updates, data drift, or workflow changes, requiring vigilance even without visible problems.

Indicators for AI Effectiveness

Recall rates, detection rates, and consistency across sites are key indicators for evaluating AI performance over time.

MMIV provides all software required for PACS integration of image AI systems as open source.
Contact us if you want to use the programs to train your AIs and integrate them into clinical workflows.