An Artificial Intelligence-based Mammography Screening Protocol for Breast Cancer: Outcome and Radiologist Workload

Authors   Lauritzen AD, Rodríguez-Ruiz A, von Euler-Chelpin MC, Lynge E, Vejborg I, Nielsen M, Karssemeijer N, Lill-holm M.

Purpose

To investigate whether an AI system could detect normal, moderate-risk, and suspicious mammograms in a screening sample to safely reduce radiologist workload and evaluate across Breast Imaging Reporting and Data System (BI-RADS) densities.

Materials and Methods

This retrospective simulation study analyzed mammographic examination data consecutively collected from January 2014 to December 2015 in the Danish Capital Region breast cancer screening program. All mammograms were scored from 0 to 10, representing the risk of malignancy, using an AI tool. During simulation, normal mammograms (score < 5) would be excluded from radiologist reading and suspicious mammograms (score > recall threshold [RT]) would be recalled. Two radiologists read the remaining mammograms. The RT was fitted using another independent cohort (same institution) by matching to the radiologist sensitivity. This protocol was further applied to each BI-RADS density. Screening outcomes were measured using the sensitivity, specificity, workload, and false-positive rate.

The AI-based screening was tested for noninferiority sensitivity compared with radiologist screening using the Farrington-Manning test. Specificities were compared using the McNemar test.

Results

The study sample comprised 114 421 screenings for breast cancer in 114 421 women, resulting in 791 screen-detected, 327 interval, and 1473 long-term cancers and 2107 false-positive screenings. The mean age of the women was 59 years ± 6 (SD). The AI-based screening sensitivity was 69.7% (779 of 1118; 95% CI: 66.9, 72.4) and was noninferior (P = .02) to the radiologist screening sensitivity of 70.8% (791 of 1118; 95% CI: 68.0, 73.5). The AI-based screening specificity was 98.6% (111 725 of 113 303; 95% CI: 98.5, 98.7), which was higher (P < .001) than the radiologist specificity of 98.1% (111 196 of 113 303; 95% CI: 98.1, 98.2). The radiologist workload was reduced by 62.6% (71 585 of 114 421), and 25.1% (529 of 2107) of false-positive screenings were avoided. Screening results were consistent across BI-RADS densities, although not significantly so for sensitivity.

Conclusion

Artificial intelligence (AI)–based screening could detect normal, moderate-risk, and suspicious mammograms in a breast cancer screening program, which may reduce the radiologist workload. AI-based screening performed consistently across breast densities.

> **Click here** to read the article

SIEMENS
Healthineers