

# Artificial Intelligence Evaluation of 122 969 Mammography Examinations from a Population-based Screening Program



**Authors** Larsen M, Aglen CF, Lee CI, Hoff SR, Lund-Hanssen H, Lång K, Nygård JF, Ursin G, Hofvind S.

## Purpose

To compare the performance of a commercially available AI system with routine, independent double reading with consensus as performed in a population-based screening program. Furthermore, the histopathologic characteristics of tumors with different AI scores were explored.

Threshold 1 was set at an AI score of 10, threshold 2 was set to yield a selection rate similar to the consensus rate (8.8%), and threshold 3 was set to yield a selection rate similar to an average individual radiologist (5.8%). Descriptive statistics were used to summarize screening outcomes.

## Materials and Methods

In this retrospective study, 122 969 screening examinations from 47 877 women performed at four screening units in BreastScreen Norway from October 2009 to December 2018 were included. The data set included 752 screen-detected cancers (6.1 per 1000 examinations) and 205 interval cancers (1.7 per 1000 examinations). Each examination had an AI score between 1 and 10, where 1 indicated low risk of breast cancer and 10 indicated high risk. Threshold 1, threshold 2, and threshold 3 were used to assess the performance of the AI system as a binary decision tool (selected vs. not selected).

## Results

A total of 653 of 752 screen-detected cancers (86.8%) and 92 of 205 interval cancers (44.9%) were given a score of 10 by the AI system (threshold 1). Using threshold 3, 80.1% of the screen-detected cancers (602 of 752) and 30.7% of the interval cancers (63 of 205) were selected. Screen-detected cancer with AI scores not selected using the thresholds had favorable histopathologic characteristics compared to those selected; opposite results were observed for interval cancer.



## Conclusion

The proportion of screen-detected cancers not selected by the artificial intelligence (AI) system at the three evaluated thresholds was less than 20%. The overall performance of the AI system was promising according to cancer detection.

> [Click here](#) to read the article

Larsen M, Aglen CF, Lee CI, Hoff SR, Lund-Hanssen H, Lång K, Nygård JF, Ursin G, Hofvind S. Artificial Intelligence Evaluation of 122 969 Mammography Examinations from a Population-based Screening Program. *Radiology*. 2022;303:502-511. doi: 10.1148/radiol.212381