Artificial intelligence in mammographic screening: Retrospective analysis of 1254 breast cancer cases from a Norwegian population-based screening program

Author/speaker: Henrik Wethe Koch (henrik.wethe.koch@sus.no / tel. 416 55 380)

Abstract

Objectives

To compare results of selected screening outcomes of a deep learning artificial intelligence (AI) system versus independent double reading in a cancer-enriched sample of screening mammograms, stratified by mammographic density.

Methods

In this retrospective study, we analyzed data from 949 screen-detected breast cancers, 305 interval cancers and 13 646 negative examinations performed in BreastScreen Norway during the period from 2010 to 2018. An AI system scored the examinations from 1-10, based on the risk of malignancy.

Results

A total of 92.7% of the screen-detected and 40.0% of the interval cancers had an AI score of 10. Sensitivity of the AI system was comparable to one of two readers in an independent double reading setting, 68.3% versus 68.6% (first reader) and 69.5% (second reader), at an equal specificity of 97.6%. Among women with the highest breast density, the AI system scored 100% of the screen-detected cancers and 48.6% of the interval cancers with an AI score of 10. Sensitivity for women with the highest breast density was 62.8% for independent double reading, and 80.9% for AI score 10. For screen-detected cancers with prior mammograms available, 41.9% had an AI score of 10 at the prior screening round.

Conclusions

The high proportion of cancers with AI score 10 and the comparable sensitivity with one reader indicates promising performance of the AI system, particularly for women with dense breasts. Results on prior mammograms with AI score 10 illustrates the potential for earlier detection of breast cancers by using AI in screen-reading.