Insufficient evidence that AI breast cancer screening is accurate enough to replace human scrutiny

September 1 2021

Credit: Pixabay/CC0 Public Domain

Humans still seem to be better than technology when it comes to the accuracy of spotting possible cases of breast cancer during screening,

suggests a review published online in *The BMJ* today.

The researchers say there is currently a lack of good quality evidence to support a policy of replacing human radiologists with artificial intelligence (AI) technology when [screening](#) for [breast cancer](#).

Breast [cancer](#) is a leading cause of death among women worldwide and many countries have introduced mammography screening programs to detect and treat it early. But examining mammograms for early signs of cancer is high volume repetitive work for radiologists, and some cancers are missed.

Previous research has suggested that AI systems outperform humans and might soon be used instead of experienced radiologists. Yet a recent review of 23 studies highlighted evidence gaps and concerns about the methods used.

To address this uncertainty, the UK National Screening Committee commissioned a team of researchers from the University of Warwick to examine the accuracy of AI for the detection of breast cancer in mammography screening practice.

The researchers reviewed 12 studies carried out since 2010 involving data for 131,822 screened women in Sweden, the United States, Germany, the Netherlands and Spain.

Overall, the quality of the methods used in the 12 studies was poor and their applicability to European or UK breast cancer screening programs was low.

Three large studies involving 79,910 women compared AI systems with the clinical decisions of the original [radiologist](#). Of these, 1,878 had screen detected cancer or interval cancer (cancer diagnosed in-between

routine screening appointments) within 12 months of screening.

The majority (34 out of 36 or 94%) of AI systems evaluated in these three studies were less accurate than a single radiologist, and all were less accurate than the consensus of two or more radiologists, which is the [standard practice](#) in Europe.

In contrast, five smaller studies involving 1,086 women reported that all of the AI systems evaluated were more accurate than a single radiologist. But the researchers note that these studies were at high risk of bias and their promising results are not replicated in larger studies.

In three studies, AI used as a pre-screen to triage which mammograms need to be examined by a radiologist and which do not screened out 53%, 45%, and 50% of women at low risk but also 10%, 4%, and 0% of cancers detected by radiologists.

The authors point to some study limitations such as excluding non-English studies that might have contained relevant evidence, and they acknowledge that AI algorithms are short lived and constantly improving, so reported assessments of AI systems might be out of date by the time of study publication.

Nevertheless, use of stringent study inclusion criteria together with rigorous and systematic evaluation of study quality suggests their conclusions are robust.

As such, they say: "Current evidence on the use of AI systems in [breast](#) cancer screening is a long way from having the quality and quantity required for its implementation into clinical practice."

They add: "Well designed comparative test accuracy studies, randomized controlled trials, and cohort studies in large screening populations are

needed which evaluate commercially available AI systems in
combination with radiologists in clinical practice."

  More information: Use of artificial intelligence for image analysis in
breast cancer screening programmes: systematic review of test accuracy,
BMJ (2021). DOI: 10.1136/bmj.n1872

Provided by British Medical Journal