Breast cancer is the most common type of cancer in women globally, occurring in about one in eight women, but early detection and treatment can considerably improve outcomes. Consequently, many developed nations have implemented large-scale mammography screening programs.

Mammography has been the frontline screening tool for breast cancer for decades with more than 200 million women being examined each year around the globe. Despite its widespread use, the interpretation of mammograms is affected by high rates of false positives and false negatives. False positives can lead to patient anxiety, unnecessary follow-up, and invasive diagnostic procedures. In contrast, a false negative can result in delayed detection, advancement of the disease, and being less amenable to treatment.

For example, in the United States, approximately 10 percent of the 40 million women who undergo breast cancer screening each year are recalled for further diagnostic imaging, and only 4-5 percent of these women are ultimately diagnosed with breast cancer. This means that more than half a million would not have to undergo unnecessary diagnostic work-up.

Another critical issue in breast cancer screening is that the workload for radiologists is high, the accuracy achieved by radiologists in cancer detection varies widely, and the performance of even the best clinicians leaves room for improvement.

Could AI Improve Screening Accuracy?

Artificial intelligence (AI)-based algorithms represent a promising avenue for concurrently improving the accuracy of digital mammography, as well as helping radiologists become more efficient by reducing this often monotonous and time-consuming task, providing them more time to focus on patient care. Moreover, breast cancer screening is perhaps an ideal application for AI in medical imaging because large curated data sets suitable for algorithm training and testing are already accessible.

With advances in the application of AI using deep learning techniques, more and more health care professionals are coming to view the technology as valuable support tools uniquely poised to help with the current challenges. A recent study performed by MIT Technology Review and GE Healthcare discovered that 79 percent of health care providers believe AI tools have helped reduce clinician burnout, allowing professionals to deliver more patient-centered, engaging care (www.technologyreview.com/hub/ai-effect/).

Given the growing interest in the use of AI in medical imaging, several newer algorithms based on deep learning have been developed for mammography.

International Evaluation of AI System

In a study published in Nature, McKinney et al evaluated the performance of a new AI system to determine if the model is capable of predicting breast cancer in mammography scans more accurately than radiologists (2020; https://doi.org/10.1038/s41586-019-1799-6).

The large international study from Google, Northwestern Medicine, and two screening centers in the United Kingdom used two large data sets of fully de-identified mammograms from the U.K. and the U.S. to train a deep-learning AI model to identify breast cancer in screening images. The test sets included scans from 23,856 women at two screening centers in the U.K. and 3,097 women at a U.S. academic medical center.

The system was then used to identify the presence of breast cancer in mammograms of women who were known to have had either biopsy-proven breast cancer or normal follow-up imaging results at least 365 days later. These predictions were then compared against the set of predictions made in clinical practice, as well as those collected from six radiologists in an independent study.

The authors provided evidence that the AI system outperformed both the historical decisions made by the radiologists who initially assessed the mammograms, and the decisions of six expert radiologists who analyzed 500 randomly selected cases in a controlled study.

The findings revealed an absolute reduction of 5.7 percent and 1.2 percent (U.S. and U.K.) in false positives and 9.4 percent and 2.7 percent in false negatives. These results held across two large data sets that are representative of different screening populations and practices.

These studies highlight the potential of AI technologies for detecting and enhancing the accuracy of breast cancer screenings when combined with assessments from human radiologists.

There are several limitations to the study. Although the U.K. dataset reflected the nationwide screening population in age and cancer prevalence, the same is not true for the U.S. dataset, which was drawn from a single screening center and enriched for cancer cases.

Additionally, apart from age, the demographics of the population studied are not well-defined. The performance of AI algorithms can be highly dependent on the population used in the training sets. Consequently, to ensure that the results are broadly applicable, it is critical that a representative sample of the general population be used in the development of this technology.
Nevertheless, the authors believe the findings of the AI system pave the way for clinical trials to improve the accuracy and efficiency of breast cancer screening.

“This is a huge advance in the potential for early cancer detection,” said study co-author Mozziyar Etemadi MD, PhD, of Northwestern Medicine in Chicago. “Breast cancer is one of the highest causes of cancer mortality in women. Finding cancer earlier means it can be smaller and easier to treat. We hope this will ultimately save a lot of lives.

“While this is exciting, validation in future trials is needed to better understand how models like these can be effectively integrated into clinical practice,” he noted.

**AI & Diagnostic Performance**

In an effort to show the benefits of concurrent use of new AI along with radiologists, a team, led by Serena Pacile, PhD, a clinical research manager in the software industry, tested MammoScreen, a tool designed to identify regions suspected of having a breast cancer on 2D mammograms and evaluate their likelihood of malignancy. The clinical investigation was recently published in *Radiology: Artificial Intelligence* (2020; https://doi.org/10.1148/ryai.2020190208).

This system produces a set of image positions with scores for suspicion of malignancy that are extracted from the four views of a standard mammogram. In this multi-reader, multi-case, retrospective study, a dataset including 240 digital mammography images, captured between 2013 and 2016, were analyzed by 14 radiologists by a counterbalance design: half were read once without AI, and the other half were read twice—one without AI and once with it. Endpoints assessed by the investigators included area under the ROC curve (area under the curve [AUC]), sensitivity, specificity, and reading time.

The findings demonstrated that overall the average AUC across readers was 0.769 (95% CI, 0.724-0.814) without the use of AI and 0.797 (95% CI, 0.754-0.840) with AI. The average difference in AUC was 0.028 (95% CI, 0.002-0.055; P=.035). The investigators said these data indicate greater interreader reliability with the aid. For 11 of 14 radiologists, MammoScreen demonstrated a trend toward lowering the false-negative rate with an average improvement of 18 percent, and the false-positive rate dropped by an average of 25 percent for eight radiologists, resulting in more standardized results.

Overall, these improvements came without adding time to a provider’s workflow, noted the researchers. For cases where the possibility of malignancy was low—less than a 2.5 percent chance—the time was about the same in the first reading session and slightly decreased in the second reading session. For those with a higher likelihood of malignancy, this time saved could lead to greater productivity, the team said.

Limitations of the study reveal the used dataset was not representative of normal screening practices. Specifically, a high rate of false-positive readings may have resulted due to readers awareness of the dataset being enriched with cancer cases, causing a laboratory effect. Furthermore, as the readers had no access to previous mammograms of the studied patients, other images, or supplementary patient information, the assessment was more difficult than a typical screening mammography reading workflow. Another factor not considered in the present study is that, in real conditions, additional factors may have an impact on reading time (i.e., stress, tiredness, etc.).

“The results show that MammoScreen may help to improve radiologists’ performance in breast cancer detection,” Pacile stated.

In March, the FDA cleared MammoScreen for use in the clinic, where it could aid in reducing the workload of radiologists. Moving forward, the team plans to investigate how this technology acts on a large screening-based population and its ability to identify breast cancers earlier.

These studies highlight the potential of AI technologies for detecting and enhancing the accuracy of breast cancer screenings when combined with assessments from human radiologists. AI technologies were able to detect pixel-level changes in tissue invisible to the human eye, while humans used forms of reasoning not available to AI. The ultimate goal will be to find the best way to combine the two to transform the future of breast imaging care for patients.

**Dibash Kumar Das is a contributing writer.**