American Journal of Roentgenology, June 2020

Should We Ignore, Follow, or Biopsy?

Impact of AI Decision Support on Breast Ultrasound Lesion Assessment · 15-Reader, 900-Case Retrospective Study

Victoria L. Mango, Mary Sun, Ralph T. Wynn, Richard Ha

The largest, most comprehensive retrospective reader study of artificial intelligence decision support in analyzing ultrasound images of breast lesions. 15 physicians each interpreted 900 cases twice in random order. All cases were pathology-proven or a minimum of 1 year follow up.

- **6 additional cancers found per 100 cases presented**
- **Up to 31% decrease in benign biopsies**
- **Reduced BI-RADS 3 follow-up recommendations**
- **Improved consistency of interpretation, both inter- and intra-operator**

> "... AI-based [decision-support] output sensitivity *and* specificity compare favorably with those of interpreting physicians ..."
>
> — Victoria Mango MD, Mammographer, Memorial Sloan Kettering

Objective

To assess the impact of artificial intelligence (AI) based decision support (DS) on breast lesion assessment.

Materials and Methods

- IRB-approved multi-center retrospective review of 900 breast lesions: 470 (52.2%) benign + 430 (47.8%) malignant
- 15 physicians (11 radiologists, 2 surgeons, 2 ob-gyns) with 3–39 years experience

- AI system (Koios DS for Breast) evaluated images and categorized as Benign, Probably Benign, Suspicious and Probably Malignant; these risk categories align with BI-RADS categories 2–4C+, respectively
- Each reader reviewed 750 cases twice: US only or US + DS; 4 weeks later, cases reviewed in opposite format
- Additionally, 150 of the 750 cases were shown again in each session. Kendall's tau-b correlation coefficient assessed intra- and inter-reader variability

Results

- Average reader US-only AUC was 0.83 (0.78 – 0.89), while average US + DS AUC was 0.87 (0.84 – 0.90) (a = 0.05). DS PLR was 1.98 (1.78, 2.18), greater than 14 of 15 readers
- 14 of 15 readers demonstrated better AUC with US + DS than US only
- Intra-reader variability improved with DS (class switching 13.6% US only vs 10.8% US + DS, p = 0.04)

CONCLUSION

AI-based DS improves accuracy of sonographic breast lesion assessment while reducing inter- and intra-operator variability.

Improved AUC UPPER GRAPH

When using Koios DS, physicians demonstrated a statistically significant improvement in AUC, improving both specificity and sensitivity.

More Accuracy, Less Variability LOWER GRAPH

The system's standalone sensitivity (98%) was higher than the average sensitivity of the 15 readers (94%).

Based on the Positive Likelihood Ratio (PLR) – the likelihood a biopsy recommendation indicates malignancy – the performance of Koios DS was at least comparable to that of fellowship-trained breast imagers.

Physicians using Koios DS improved, reducing both inter- and intra-reader variability. In other words, they agreed with each other and with themselves more often when using the system. Koios DS decreased the variability and gap in performance between specialists and non-specialists, promising to elevate the quality of care at centers where not all breast imagers are fellowship-trained.

Subsets of readers demonstrated improvements in weak areas. For example, highly specific but less sensitive readers saw greater improvements in sensitivity without compromising specificity.

Works Across Vendors and Frequencies

Koios DS Breast demonstrated no statistical difference in DS system performance across the various ultrasound transducer frequencies and equipment manufacturers tested. The study included images from multiple institutions.

> "Given the performance of the standalone system, the [decision-support] output may have a larger impact if it is used more frequently."