

## Title: High Performance of FDA-Cleared Platform for Mammography Triage

**Authors:** Tara A. Retson, MD, PhD<sup>1</sup>, Vivian Lim, MD<sup>1</sup>, Alyssa T. Watanabe, MD<sup>2</sup>

<sup>1</sup>UCSD Department of Radiology, La Jolla, CA, <sup>2</sup>USC Keck School of Medicine, Los Angeles, CA

**Background:** Compared to computer aided detection (CAD) programs which highlight individual imaging features, triage programs prioritize or flag exams within a radiology worklist. Traditional CAD has not been shown to significantly benefit cancer detection in screening mammography, however, recent studies suggest that artificial intelligence (AI) based triage programs could improve cancer detection and expedite radiologist workflow (1,2). Here, we sought to evaluate the performance of a commercial AI-based triage algorithm on exams with varying breast densities and lesion types.

**Methods:** This retrospective, multi-center, and multi-vendor study examined 1255 screening mammograms consisting of 4 standard views (LCC, LMLO, RCC, RMLO). The population was enriched with biopsy-confirmed cancers, containing 400 positive and 855 normal (BIRADS 1 and 2) studies. Images were analyzed by a commercially available AI algorithm (cmTriage, Curemetrix) and given a quantitative score based on suspicion for cancer or recall. Triage was then performed at a study level where exams were labeled “suspicious,” or left unlabeled (indicating a low suspicion exam).

**Results:** The algorithm demonstrated an area under the curve (AUC) of 0.95 (95% CI: 0.94 - 0.96) for lesion identification. This AUC held across densities (0.95) and lesion types (masses: 0.94, 95% CI: 0.92 to 0.96; or microcalcifications: 0.97, 95% CI: 0.96 to 0.99). While the algorithm has a default specificity of 93% (modifiable up to 99%), to evaluate real world performance we used 86.9% sensitivity (95% CI: 83.6% to 90.2%), as was observed for practicing radiologists by the Breast Cancer Surveillance Consortium (BCSC) study (3). Resulting algorithm specificity was 88.5% (95% CI: 86.4% to 90.7%), similar to BCSC radiologist specificity of 88.9%, and indicating that algorithm performance may be comparable to real-world practice. Average study turnaround was 3.35 minutes at 10Mbps/s upload and 37Mbps/s download, within the clinical operational expectations of breast cancer screening.

**Conclusions:** AI-based triage software can perform at and above the level of practicing radiologists. By drawing attention to suspicious exams, AI-based triage may provide positive reader bias to improve accuracy, and potentially act as a second reader particularly for negative exams. As a workflow improvement, it could enable faster recall and immediate patient notification of low suspicion studies, reducing patient stress and improving care.

### References:

1. Yala A, Schuster T, Miles R, Barzilay R, Lehman C: A Deep Learning Model to Triage Screening Mammograms: A Simulation Study. *Radiology* 2019; 293:38–46.
2. Dembrower K, Wåhlin E, Liu Y, et al.: Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Heal* 2020; 2:e468–e474.
3. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL: Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015; 175:1828–1837.