
MammosighTR: Nationwide BI-RADS-Labelled Mammograms



Large, well-annotated mammography resources are essential for developing and evaluating artificial intelligence in population screening, particularly when labels include Breast Imaging Reporting and Data System (BI-RADS) categories, breast composition and quadrant-level localisation. Such structure supports both classification and localisation tasks while reflecting real-world variability in participants, equipment and image quality control. In this context, a nationwide, centrally reported screening resource from Türkiye provides BI-RADS labels with detailed descriptors drawn from double-read activity across 347 centres.

A blinded, multitier radiologist workflow underpins annotations, and an explicit enrichment strategy counterbalances the rarity of suspicious and malignant findings in routine screening, with clear guidance for reweighting when estimating population-level metrics. Agreement analyses and external benchmarks characterise task difficulty across BI-RADS strata, and access planning aligns with national data governance.

Nationwide Screening and Curated Case Mix

Biennial mammography has been offered to women aged 40–69 years with centralised digital reporting since 2016. By 1 January 2023 the national screening pool comprised 3,511,025 cases. Between March 2016 and December 2022, distribution in the pool was 93.6% BI-RADS 1–2, 5.11% BI-RADS 0, 0.43% BI-RADS 4, and 0.24% BI-RADS 5. From this pool, 14,352 cases were sampled for review and 12,740 complete cases were retained after exclusions for incomplete views, poor image quality or participant identifiers, implants were excluded to avoid biasing BI-RADS assessment. The case corresponds to one screening episode with standard mediolateral oblique and craniocaudal projections for both breasts in DICOM format.

Must Read: [MBI with DBT Raises Invasive Cancer Detection in Dense Breasts](#)

The image corpus spans multiple vendor systems, most commonly FUJIFILM (91%, 11,591/12,740), followed by SIEMENS (6%, 774/12,740), GIOTTO (1.3%, 165/12,740) and other vendors (1.6%, 209/12,740). Acquisition types were 89.1% digital mammography (11,357/12,740) and 10.9% computed radiography (1,382/12,740). Participant ages included 5,636 aged 40–49 years, 4,575 aged 50–59 years, 2,489 aged 60–69 years and 40 aged 70–79 years. To ensure adequate representation of clinically significant but infrequent lesions, the dataset was enriched with biopsy-confirmed suspicious or malignant cases, resulting in 39.4% BI-RADS 4–5 (5,021/12,740) in the curated set. Detailed sampling proportions are provided to enable weighting when estimating population metrics. BI-RADS 1–2 reflect benign outcomes after follow-up and BI-RADS 0 preserves uncertainty where imaging or clinical evidence did not resolve categorisation.

Multitier Radiologist Labelling and Scope

Nineteen radiologists with 5–15 years of experience contributed labels through a blinded, multitier workflow. Two primary readers first assigned structured labels while blinded to prior BI-RADS and histopathology. Six senior radiologists validated these assignments, also blinded to prior labels and pathology. A final senior reviewer resolved disagreements and allocated cases to training and test sets. Radiologist consensus served as the reference standard, with histopathology and follow-up informing selection of high-certainty cases rather than dictating labels.

Each case includes breast composition (categories a–d), BI-RADS category spanning 0, 1–2 and 4–5, and quadrant-level localisation specifying the region driving the assigned BI-RADS category. Where multiple findings existed, the highest-risk BI-RADS label was recorded. Quadrant localisation covers upper-outer, upper-inner, lower-outer, lower-inner and central regions. Use of a structured national teleradiology interface standardised reporting of composition, quadrant and BI-RADS across sites and devices. Within enriched BI-RADS 4–5 cases, heterogeneously dense breasts (composition c) were most common. Potential sources of bias are acknowledged, notably oversampling of BI-RADS 4–5 and varying verification across categories. Recommended mitigations include reweighting for population estimates and class-aware learning

strategies that maintain sensitivity to rare classes.

Agreement, Benchmarks and Access

Interreader agreement was measured on a 1,000-case subset assembled with a dual-strata design that preserved real-world prevalence while enriching diagnostically challenging cases such as BI-RADS 0, dense breasts and central lesions. Agreement was high for BI-RADS category (Fleiss' $\kappa = 0.85$), breast density ($\kappa = 0.74$) and quadrant localisation ($\kappa = 0.81$). Lower agreement appeared in heterogeneously dense breasts ($\kappa = 0.68$), BI-RADS 0 ($\kappa = 0.72$) and centrally located lesions ($\kappa = 0.69$), reflecting recognised areas of ambiguity. A summary table reports agreement percentages across localisations including upper-outer, upper-inner, lower-outer, lower-inner, central and multiple-quadrant regions.

The resource supported an Artificial Intelligence in Healthcare competition at TEKNOFEST 2023 with 42 finalist teams. Mean F1 scores increased from BI-RADS 0 (0.255, 95% CI 0.207–0.304) to BI-RADS 1–2 (0.417, 95% CI 0.340–0.493) and BI-RADS 4–5 (0.537, 95% CI 0.427–0.647). Additional metrics summarised mean specificity, sensitivity and accuracy with confidence intervals by category, illustrating variability in performance, particularly for uncertain and malignant strata. Correlation analysis showed a strong positive association between team F1 scores for BI-RADS 1–2 and 4–5.

Construction followed permission from the Ministry of Health with full anonymisation under national data protection regulations that distinguish anonymous, anonymised and pseudonymised data. Legal provisions allow processing and sharing of health data for public health, service planning and scientific research where individual rights and national interests are protected. Public release is planned via the official open data repository upon acceptance and e-publication in line with national open data policy.

A nationally derived, carefully curated mammography dataset with BI-RADS categories, breast composition and quadrant-level localisation offers a substantive foundation for AI research in breast screening. Vendor diversity, standardised double-reading and a blinded, multitier review provide a robust labelling framework, while explicit enrichment and guidance for reweighting support sound generalisability. Agreement analysis and competition benchmarks indicate expected difficulty by BI-RADS strata. With planned open release under national data governance, the resource is positioned to support evaluation and potential efficiency gains in screening workflows.

Source: [Radiology: Artificial Intelligence](#)

Image Credit: [iStock](#)

Published on : Sat, 11 Oct 2025