
Open-Source AI for Breast Density Grading



Breast density influences both cancer risk and interpretive accuracy in mammography, yet routine assessments remain variable across readers. A new open-source approach aims to bring consistency by combining a custom convolutional neural network with an extreme learning machine layer to classify density into BI-RADS categories A–D. Trained on more than ten thousand full-field digital mammograms with double reading and consensus adjudication, the system reports high agreement with specialist ratings and provides a free, deployment-ready tool designed for resource-limited settings. External validation on a long-standing public dataset offers an additional view of performance and generalisability, highlighting strengths and areas for refinement in borderline classes.

Dataset and Clinical Grounding

The retrospective database comprised 10,371 mammographic images from 2,472 patients, with a mean age of 55.2 years. Most patients contributed four images, though some had two following mastectomies. Images were independently reviewed by at least two board-certified breast imaging specialists, with discrepancies resolved by consensus under BI-RADS A–D density criteria. Approximately 55% of breasts were A or B, and 45% were C or D. The project received ethics approval and used k-fold cross-validation that kept all images from a patient within a single fold to avoid leakage.

Double reading was part of routine care and provided a robust reference for training and evaluation. Inter-rater agreement between the two specialists was nearly perfect, supporting the reliability of the ground truth against which the model was assessed. The workflow focused on images with acceptable positioning and a clearly reported density, tightening the clinical relevance of the dataset while minimising label ambiguity. These controls underpin the reported metrics and situate the tool in real-world interpretation conditions where density grading is known to vary.

Model Design and Evaluation

The authors explored both custom models and established architectures such as VGG, ResNet, MobileNet and DenseNet, implementing transfer learning where appropriate. The final system—referred to as model 3 plus an extreme learning machine layer—integrated four convolutional layers, four max-pooling layers, three fully connected layers and dropout, paired with categorical cross-entropy and the Adam optimiser. Across systematic tests of resolution, iterations, batch sizes, hidden units and k-fold settings, the optimal configuration used 128×128 inputs, 300 iterations, a batch size of 32 and 128 hidden units in the ELM. This balance optimised accuracy and computation time.

Must Read: [AI-Guided Hybrid Reading Maintains Mammography Accuracy](#)

Under 10-fold cross-validation, the model achieved an overall testing accuracy of 95.4%, specificity of 98.0% and sensitivity of 92.5%. Per-class performance showed high precision and F1 scores, with slightly lower recall for category C relative to A and B, indicating a known challenge in separating heterogeneously dense tissue from adjacent classes. Agreement between the automated output and specialist consensus reached a weighted kappa of 0.90, close to the near-perfect inter-specialist kappa of 0.95. These figures indicate strong concordance with experienced readers while signalling room for refinement in mid-range density distinctions.

External Performance and Practical Access

To assess generalisability, the system was tested on the Mini-MIAS database, which labels density as fatty, fatty-glandular or dense-glandular rather than BI-RADS A–D. The authors mapped BI-RADS outputs to these three classes to enable comparison. On this independent dataset, the model yielded 73.9% accuracy, 81.1% precision, 87.3% specificity and 75.1% sensitivity. Confusion patterns suggested stronger performance on

clear fatty or dense-glandular cases, with more ambiguity in the transitional fatty-glandular category, aligning with the internal observation that category C distinctions are comparatively harder.

The system is made available as an open-source tool with a Streamlit interface that accepts DICOM, PNG, JPG, or JPEG. Images are resized and normalised prior to classification, returning a BI-RADS density category designed to complement clinical workflows. By disclosing code and providing an accessible application, the authors target reproducibility and low-barrier deployment, addressing a gap in a field where many commercial or academic solutions remain proprietary or require significant infrastructure.

A freely accessible AI system for breast density classification demonstrates high concordance with specialist assessment on a large, clinically grounded dataset and shows competitive performance on an independent public archive after category mapping. Internal metrics indicate strong accuracy and specificity, with an identifiable challenge around heterogeneously dense cases that mirror clinical experience. The combination of open code, a usable interface and patient-level cross-validation supports practical adoption, particularly in settings where commercial tools are cost-prohibitive. For imaging departments and screening programmes, the approach offers a scalable option to reduce variability in density reporting and to support more consistent risk communication and follow-up decisions.

Source: [Journal of Imaging Informatics in Medicine](#)

Image Credit: [iStock](#)

Published on : Sun, 21 Sep 2025