

Towards Explainable Techniques for Automatic Segmentation in Breast Density Analysis

Oscar Stanchi^{1,2,3,4} , Juan Seery¹ , Facundo Quiroga^{1,4,5} , Franco Ronchetti^{1,5} , Domenec Puig³ , and Hatem Rashwan³ 

¹ III-LIDI, Facultad de Informática, UNLP, La Plata, Argentina

² CONICET, Argentina

³ DEIM, Universitat Rovira i Virgili, Tarragona, España

⁴ Maestría en Data Mining, FCEN, Universidad de Buenos Aires, Buenos Aires, Argentina

⁵ CIC-PBA, Argentina

{ostanchi,jmseery,fquiroga,fronchetti}@lidi.info.unlp.edu.ar,
{domenec.puig,hatem.abdellatif}@urv.cat

Abstract. Breast density is a clinically relevant imaging biomarker, yet dense-tissue segmentation is rarely available at scale because pixel-level annotation is time-consuming and intrinsically ambiguous. We propose a weakly supervised framework that leverages image-level BI-RADS density labels to produce pixel-wise dense-tissue masks in mammograms, bypassing manual annotation. The framework generates class-conditional saliency maps via multi-stage Grad-CAM aggregation with Top- $t\%$ pooling and derives binary masks through Otsu thresholding with anatomical masking. We curate a cohort of 4,387 mamm

Hacia Técnicas Explicables para la Segmentación Automática en el Análisis de Densidad Mamaria

Abstract. La densidad mamaria es un biomarcador de imagen clínicamente relevante; sin embargo, la segmentación del tejido denso rara vez está disponible a gran escala debido a que la anotación a nivel de píxel es costosa e intrínsecamente ambigua. Proponemos un marco de segmentación débilmente supervisado que utiliza etiquetas de densidad BI-RADS a nivel de imagen para producir máscaras de tejido denso a nivel de píxel en mamografías, prescindiendo de la anotación manual. El marco genera mapas de saliencia condicionales por clase mediante agregación multi-etapa de Grad-CAM con agrupamiento Top- $t\%$, y obtiene máscaras binarias mediante umbralización de Otsu con enmascaramiento anatómico. Se curó una cohorte de 4.387 mamografías del dataset RSNA Screening Mammography Breast Cancer Detection, con un subconjunto de 100 imágenes anotadas por radiólogos para evaluación. Bajo un protocolo que entrena sobre categorías extremas (BI-RADS A vs. D) y evalúa sobre A vs. D^* ($D^* = \{B, C, D\}$), una configuración preliminar con ResNet-152 alcanza un índice Dice de $0,694 \pm 0,018$ (desviación estándar entre semillas sobre $k = 30$ corridas), superando sustancialmente la referencia clásica LIBRA (0,405). Estos resultados demuestran que la saliencia débilmente supervisada puede producir delineaciones de tejido denso clínicamente plausibles.

Keywords: densidad mamaria , mamografía , aprendizaje débilmente supervisado , mapas de saliencia , segmentación de imágenes médicas

1 Introducción

La densidad mamaria es un factor de riesgo independiente para el cáncer de mama y un determinante clave de la sensibilidad mamográfica: el tejido fibroglandular denso puede enmascarar lesiones malignas, motivando legislación que exige su notificación (Chen et al., 2025; Spak et al., 2017). Los radiólogos evalúan la densidad en la escala BI-RADS de cuatro categorías—A (predominantemente graso) a D (extremadamente denso)—pero esta valoración es subjetiva y no proporciona información espacial sobre *dónde* se localiza el tejido denso (Spak et al., 2017).

Los métodos existentes presentan limitaciones complementarias. Las herramientas clásicas como LIBRA (Keller et al., 2012) segmentan el tejido denso mediante fuzzy c-means y SVM, pero su precisión es limitada (Gudhe et al., 2022). Los clasificadores basados en deep learning alcanzan precisión a nivel de radiólogo para categorías BI-RADS (Chen et al., 2025), pero producen solo una etiqueta global. Los métodos débilmente supervisados como GLAM (Liu et al., 2021) y GMIC (Shen et al., 2021) utilizan CAM (Zhou et al., 2016) o Grad-CAM (Selvaraju et al., 2017) para localizar regiones discriminativas, pero fueron diseñados para lesiones focales, no para el patrón difuso del tejido denso. La localización espacial del tejido denso difuso mediante supervisión débil permanece, por tanto, poco explorada.

La segmentación completamente supervisada requeriría máscaras a nivel de píxel, que son escasas e inviables a la resolución nativa de las mamografías (Agbodike et al., 2025; Behravan et al., 2024). Proponemos un marco débilmente supervisado que aprovecha etiquetas BI-RADS a nivel de imagen para producir máscaras de tejido denso sin supervisión de segmentación. Las contribuciones preliminares son: (i) un dataset curado de 4.387 mamografías del RSNA (Carr et al., 2022) con 100 máscaras expertas; (ii) un pipeline basado en saliencia multi-etapa Grad-CAM con Top- t % pooling; y (iii) un protocolo de evaluación A vs. D / A vs. D* que evalúa la generalización a densidades intermedias.

2 Metodología

2.1 Dataset

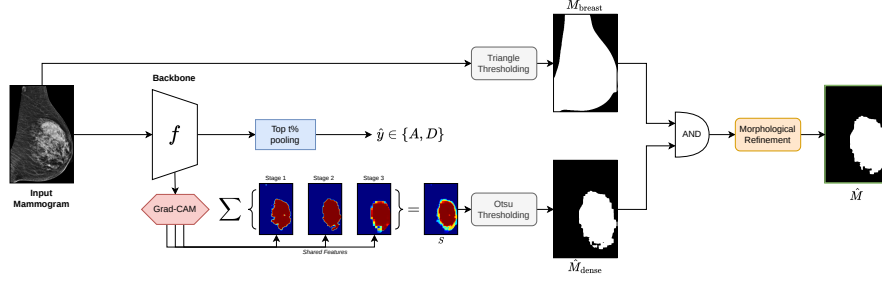


Fig. 1: Pipeline propuesto: supervisión a nivel de imagen, generación de mapas de saliencia multi-etapa y conversión a máscaras de tejido denso.

1×1 por etapa proyecta las características intermedias a C canales (uno por clase BI-RADS), produciendo tensores de saliencia condicionales por clase.

Top- $t\%$ pooling. Sea $\mathbf{S}^{(l)} \in \mathbb{R}^{C \times H \times W}$ el tensor de saliencia en la etapa $l \in \{1, 2, 3\}$. El logit de clase se obtiene promediando las $K = \lfloor t \cdot H \cdot W \rfloor$ activaciones más altas. La predicción global agrega las tres etapas: $\bar{y}_c = \frac{1}{3} \sum_l y_c^{(l)}$. Este pooling incentiva una respuesta espacialmente distribuida, a diferencia del max pooling (un único pico) o del GAP (señal diluida). El parámetro t se exploró en $[0, 1, 0, 5]$, seleccionándose $t = 0, 3$ por validación sobre un subconjunto del entrenamiento.

Saliencia y conversión a máscara. Se estima sal

Table 1: Resultados preliminares (protocolo A vs. D*, 100 imágenes anotadas). Dice, IoU y W-F1: media $\pm$ desviación estándar entre semillas ($k = 30$ corridas); LIBRA reporta solo la media (sin barrido de semillas).

Método	Dice $\uparrow$	IoU $\uparrow$	Test W-F1 $\uparrow$
ResNet-152 (propuesto)	$0,694 \pm 0,018$	$0,577 \pm 0,017$	$0,90 \pm 0,009$
LIBRA	0,405	0,297	—

3 Resultados Preliminares

La Tabla 1 presenta los resultados preliminares con ResNet-152 como backbone bajo el protocolo A vs. D*, evaluados sobre las 100 imágenes anotadas.

ResNet-152 alcanza un índice Dice de $0,694 \pm 0,018$ y un IoU de $0,577 \pm 0,017$, con un F1 ponderado de 0,90, superando a LIBRA (Dice 0,405; IoU 0,297) por un margen considerable.

El análisis por categoría (índice Dice medio por clase) revela que BI-RADS C (0,654) y D (0,718) mantienen rendimiento compacto y estable. BI-RADS A presenta un patrón bimodal (media 0,880): máscara vacía correcta en la mayoría de casos, pero falsos positivos en 3/25 imágenes. La mayor debilidad se observa en BI-RADS B (0,525), esperable dado que esta categoría intermedia no participa del entrenamiento. Los falsos positivos en A y la sobreestimación en B sugieren que el modelo tiende a sobreestimar la densidad, un sesgo coherente con el entrenamiento exclusivo en extremos. Las categorías clínicamente más relevantes (C y D) mantienen rendimiento estable, sugiriendo que el marco propuesto es

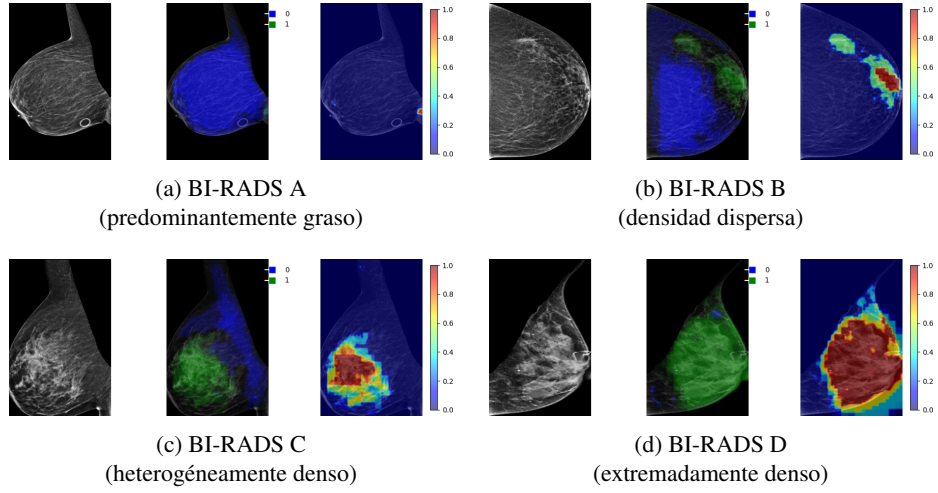


Fig. 2: Resultados cualitativos por categoría BI-RADS. Cada panel muestra: mamografía original, máscara predicha y mapa de saliencia.

capacidad representacional sobre la calidad de segmentación; (ii) entrenamiento con las cuatro categorías BI-RADS con diseño de pérdida apropiado para el ruido de etiquetas en las categorías intermedias; (iii) incorporación de modelos fundacionales (e.g., SAM) para refinamiento de bordes; (iv) fusión multi-vista (CC + MLO); y (v) validación en datasets externos como VinDr-Mammo (Nguyen et al., 2023).

Bibliography

- Agbodike, O., Kuo, C.-F., and Chen, J. (2025). Deep multi-head CNN for unified breast density segmentation and BI-RADS estimation in 2D mammograms. In *Medical Imaging with Deep Learning – Short Papers*.
- Behravan, H., Gudhe, N. R., Okuma, H., Sudah, M., and Mannermaa, A. (2024). A dataset of mammography images with area-based breast density values, breast area, and dense tissue segmentation masks. *Data in Brief*, 57:110980.
- Carr, C., Kitamura, F., Partridge, G., et al. (2022). RSNA screening mammography breast cancer detection. <https://kaggle.com/competitions/rsna-breast-cancer-detection>. Kaggle.
- Chen, C., Wang, E., Wang, V. Y., Chen, X., Feng, B., Yan, R., Zhu, L., and Xu, D. (2025). Deep learning prediction of mammographic breast density using screening data. *Scientific Reports*, 15(1):11602.
- Fan, H., Xie, F., Li, Y., Jiang, Z., and Liu, J. (2017). Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. *Computers in Biology and Medicine*, 85:75–85.
- Gudhe, N. R., Behravan, H., Sudah, M., Okuma, H., Vanninen, R., Kosma, V.-M., and Mannermaa, A. (2022). Area-based breast percentage density estimation in mammograms using weight-adaptive multitask learning. *Scientific Reports*, 12(1):12060.
- Keller, B. M., Nathan, D. L., Wang, Y., Zheng, Y., Gee, J. C., Conant, E. F., and Kontos, D. (2012). Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Medical Physics*, 39(8):4903–4917.
- Liu, K., Shen, Y., Wu, N., Chłędowski, J., Fernandez-Granda, C., and Geras, K. J. (2021). Weakly-supervised high-resolution segmentation of mammography images for breast cancer diagnosis. *Proceedings of Machine Learning Research*, 143:451–472.
- Nguyen, H. T., Nguyen, H. Q., Pham, H. H., Lam, K., Le, L. T., Dao, M., and Vu, V. (2023). VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data*, 10(1):277.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh

- Uyun, S., Hartati, S., Harjoko, A., and Choridah, L. (2015). A comparative study of thresholding algorithms on breast area and fibroglandular tissue. *International Journal of Advanced Computer Science and Applications*, 6(1):120–124.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2929.