A Deep Learning Classifier Using Sliding Patches for Detection of Mammographical Findings

Diego Mellado, Marvin Querales, Julio Sotelo, Eduardo Godoy, Fabian Pardo, Scarlett Lever, Stéren Chabert, Rodrigo Salas
Doctorado en Ciencias e Ingeniería para la Salud, Universidad de Valparaíso

November 16th, 2023

Universidad de Valparaíso CHILE

This work was funded by:

- National Agency for Research and Development (ANID) / Scholarship program / *Becas Doctorado Nacional 2022* - 21221429
- Millenium Science Initiative Program ICN2021_004
- ANID FONDEF IDEA I+D ID20i10332 “*Artificial Intelligence System for Support in Diagnosis and Priorization of Mammographic Exams*”
- ANID FONDECYT Research Grant 1221938

ihealth | Millennium Institute for Intelligent Healthcare Engineering

Agencia Nacional de Investigación y Desarrollo
Ministerio de Ciencia, Tecnología, Conocimiento e Innovación
Gobierno de Chile

Our study proposes a Deep Learning Classifier trained on cropped segments of mammograms for multi-label classification of pathological findings.
General Inspection of the complete image using local patches.

Dataset

- **Vindr-mammo** Dataset [1]
- 5000 digital mammogram examinations containing for each view:
  - BIRADs
  - Breast density
  - Bounding box of each finding present on the image
- 10 classes of pathological findings

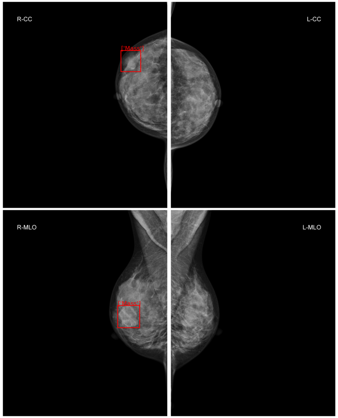

Figure: Example of a mammogram from the Vindr-mammo dataset.

[1] Hieu Huy Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. *VinDr-Mammo: A Large-Scale Benchmark Dataset for Computer-Aided Detection and Diagnosis in Full-Field Digital Mammography*. Version 1.0.0. PhysioNet. DOI: 10.13026/BR2V-7517. URL: https://physionet.org/content/vindr-mammo/1.0.0/ (visited on 11/08/2022)

Dataset Distribution

- Each DICOM Image is normalised to a range of $[0, 1]$
- Apply CLAHE algorithm [2] for histogram equalization at 2 different scales
- Fuse the equalized images with the original, channel-wise
- Image is cropped to the bounding box of the breast using Otsu's thresholding.

[2] S.M. Pizer, R.E. Johnston, J.P. Ericksen, et al. "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness". In: *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*. [1990] Proceedings of the First Conference on Visualization in Biomedical Computing. May 1990, pp. 337–345. DOI: 10.1109/VBC.1990.109340. URL: https://ieeexplore.ieee.org/document/109340 (visited on 10/03/2023)

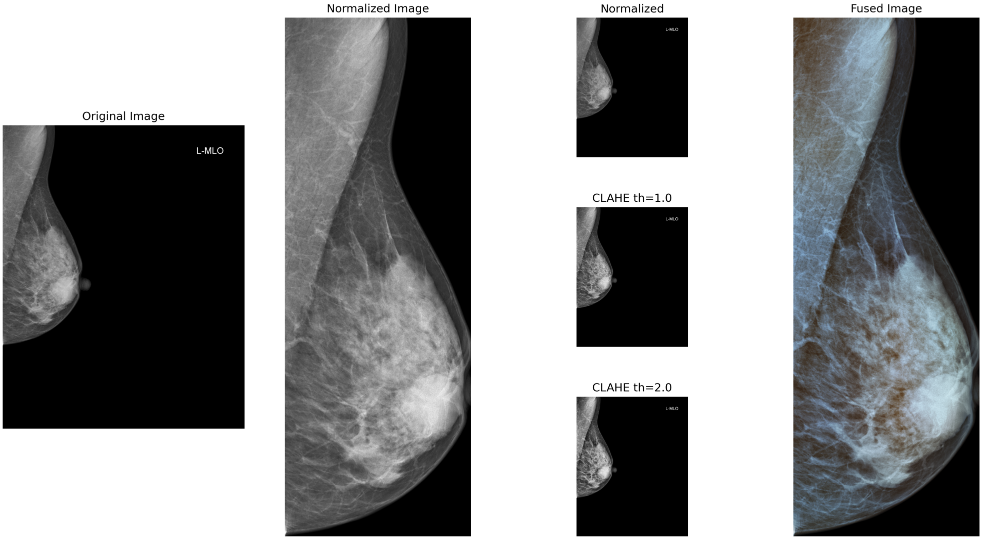

Figure: Image processing pipeline

Evaluation of deep learning architectures

We evaluated the following architectures as a backbone:

- ResNet50
- EfficientNetV2
- DenseNet
- Swin Transformer
- MobileNet
- VGG19

Classifier Layer using 2 layer MLP with 512 hidden units and ReLU activation. Output layer with 10 units and sigmoid activation.

- We use the backbone of the selected architecture as a feature extractor.
- The output of the backbone is flattened and fed to the classifier.
- The classifier is trained using the cropped images.

- For each finding present on an image, we sampled a random area of the finding.
- Area and aspect ratio randomly sampled from an Uniform distribution.
  - Area: [.005, 5.] of the finding bounding box area.
  - Aspect ratio: [0.331.66] of the finding bounding box aspect ratio.
  - The bounding box center is uniform sampled from 50 px around the original center.
- For normal images (*No Finding*), we sample a random area of the breast, with similar sampling parameters.
- Sampled image is resized to 256×256 pixels.
- **image augmentation**: Contrast, brightness, saturation, flip and rotation are randomly modified.

- Focal Loss Function

$$\mathcal{L}_{\text{Focal}} = -(1 - p_t)^{\gamma} \log(p_t)$$

- Adam Optimizer with a starting learning rate of 1×10^{-4}
- Reduce on plateau with a patience of 5 epochs and a factor of 0.1
- Weighted sampling of the dataset, in order to reduce imbalance.

Local Information

- Segmentation of the complete image into a set of overlapping windows of size 256×256 pixels and stride of 32 px.
- Reconstruction of prediction heatmap from the overlapping windows.

Table: F1 Score for Pathological findings classification task with Vindr, comparing different Deep Learning Models.

	N	DenseNet	EfficientNet	ResNet50	SwinTransformer	VGG19	MobileNet
Mass	237	0.783	0.815	0.742	0.770	0.756	0.708
Suspicious Calcification	115	0.847	0.865	0.860	0.828	0.873	0.828
Assymetries	79	0.306	0.295	0.200	0.310	0.204	0.324
Suspicious Lymphonode	11	0.667	0.500	0.737	0.370	0.400	0.476
Weighted Average	442	0.712	**0.727**	0.675	0.693	0.679	0.665

EfficientNet has the best overall performance, with an F1 score of 0.727. We chose this model as the backbone for our classifier.

Table: Metrics of pathological findings classification task using Vindr

	Accuracy	Precision	Recall	F1	Support
No Finding	0.978	0.991	0.984	0.988	3643
Mass	0.970	0.743	0.730	**0.736**	237
Suspicious Calcification	0.990	0.805	0.861	**0.832**	115
Focal Asymmetry	0.983	0.200	0.094	0.128	53
Architectural Distortion	0.993	0.222	0.083	0.121	24
Asymmetry	0.995	0.000	0.000	0.000	20
Suspicious Lymph Node	0.998	0.625	0.455	0.526	11
Skin Thickening	0.998	1.000	0.333	0.500	12
Nipple Retraction	0.998	0.000	0.000	0.000	7
Global Asymmetry	0.998	0.000	0.000	0.000	6
Skin Retraction	0.999	0.500	0.333	0.400	3

Our model achieves a weighted Accuracy of 93.80 % and a weighted F1 score of 0.9557

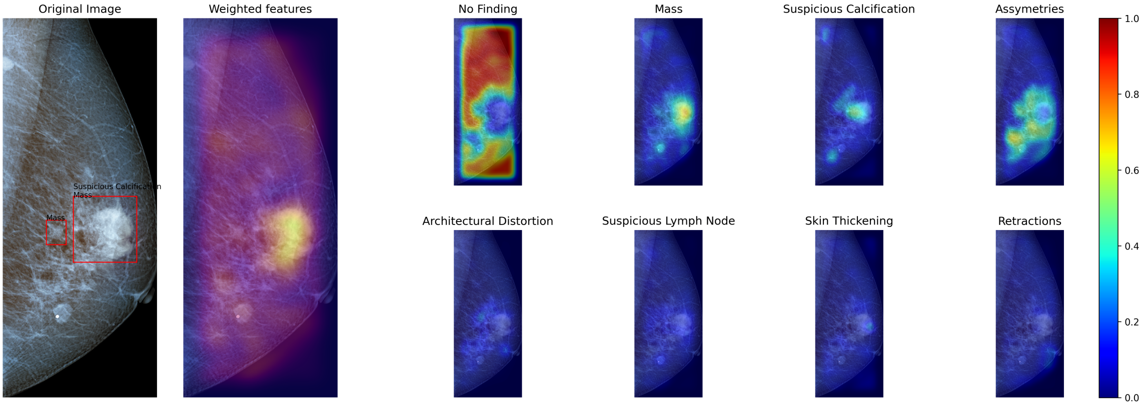

Figure: Cropped test image with bounding boxes of its findings, feature activation heatmap, and prediction of the classifier.

Discussion