Deep Learning Explainability for Breast Cancer Detection in Mammography

Karla Sam, Robert Martí
ViCOROB Institute, University of Girona, Girona, Spain

1. INTRODUCTION

Recent concerns over the need to understand the decision-making process of DL models has generated the development of several explainable artificial intelligence (XAI) methods. The purpose of this work was to apply XAI techniques to two models for two different tasks: patch-based breast cancer classification and whole mammogram classification. The resulting attribution maps from the different explainability methods were qualitatively and quantitatively assessed.

2. DATASET

		Train	Validation	Total
OMIDB Subset (Iceberg Selection)	**Non-malignant**	3045	763	3808
	Malignant	3045	763	3808
	Total	6090	1526	**7616**
RSNA22 Subset	**Non-malignant**	1647	411	2058
	Malignant	562	147	709
	Total	2209	558	**2767**
OMIDB Hologic Subset (Full-Images)	**Non-malignant**	2896	719	3615
	Malignant	2892	722	3614
	Total	5788	1441	**7229**

3. METHODS

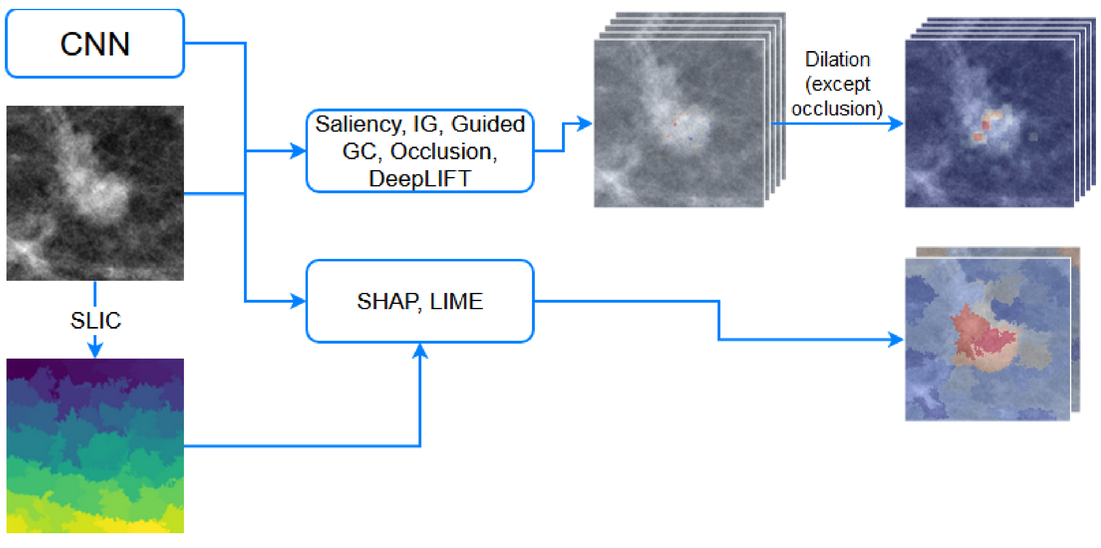

Fig. 1. Diagram showing the generation of the heatmaps from the XAI methods. The model and the input image are entered into the methods. SHAP and LIME, additionally, take the segmented image. The resulting heatmaps are dilated for better visualization (except for Occlusion, SHAP, and LIME).

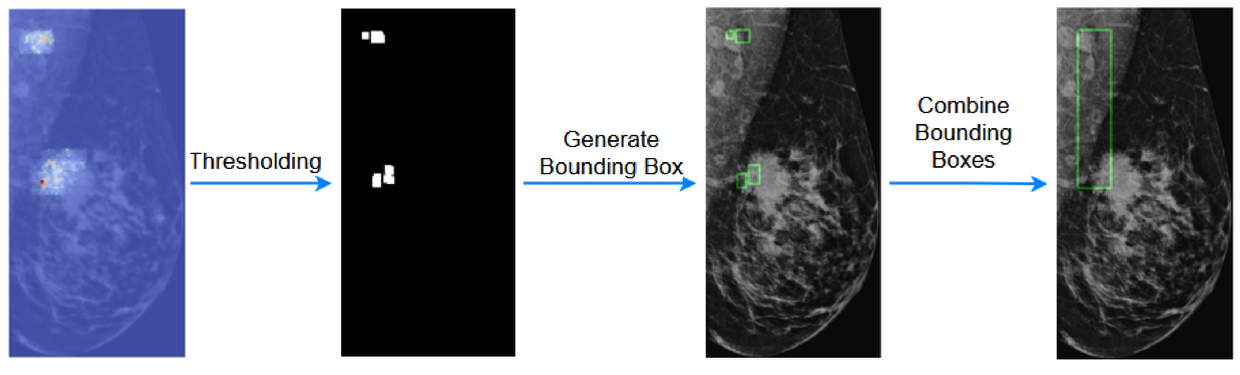

Fig. 2. Diagram of the general process for obtaining the bounding box per whole mammogram. The dilated attribution maps are binarized above the 95th percentile. The bounding boxes are obtained per contour found in the binarized image, and, in case of multiple bounding boxes, they are combined into a single bounding box.

RESULTS

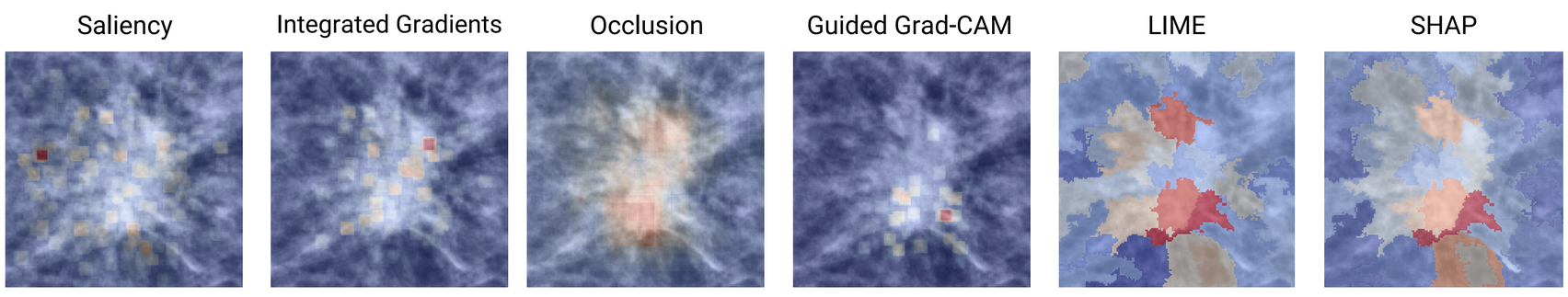

Fig. 3. Sample results of the heatmaps from the various explainability methods.

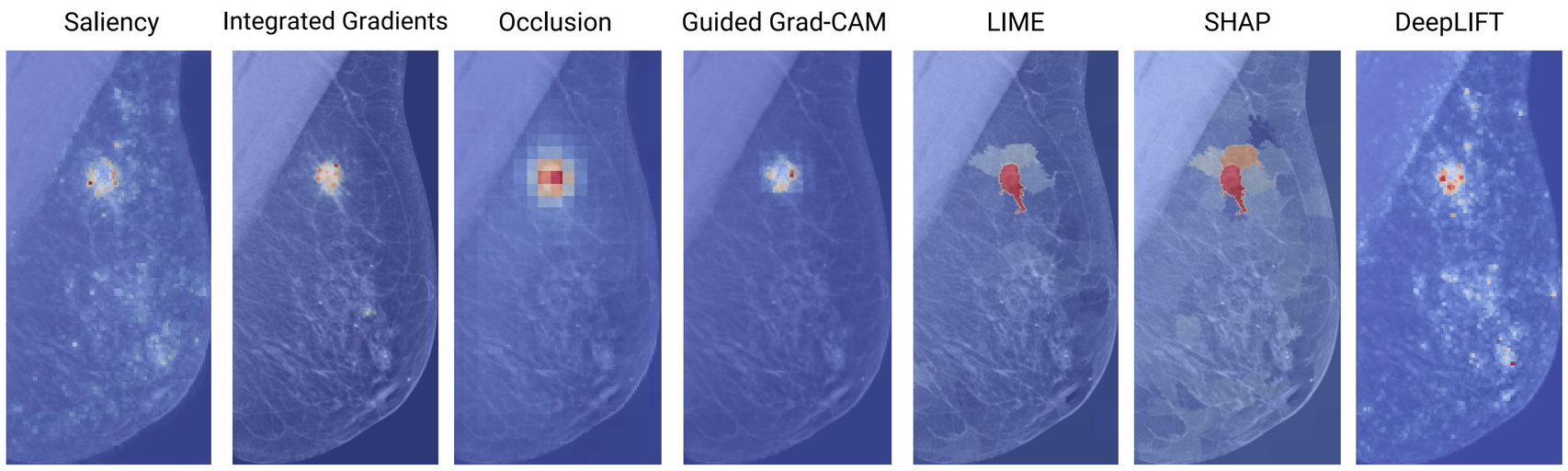

Fig. 4. Sample results of the heatmaps from the different explainability methods.

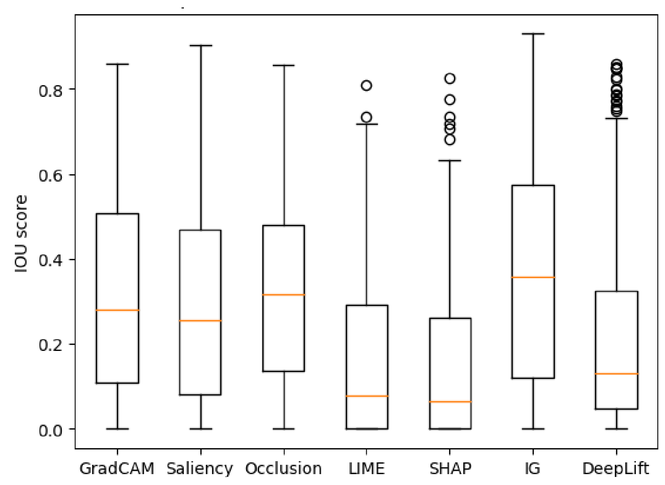

Fig. 5. Box plots for the intersection over union (IOU) scores for each XAI method for all of the full images.

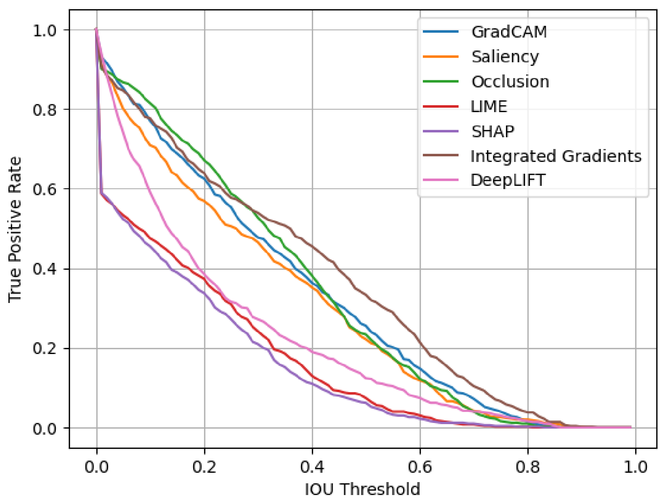

Fig. 6. IOU threshold vs True Positive Rate graph for each XAI algorithm.

CONCLUSIONS

The results from the XAI techniques indicate that they provide useful information for understanding the decision-making processes of a network in medical imaging. When correctly trained, the methods highlight the areas with clinical relevance, which in this case correspond to the lesions in malignant mammograms. Integrated Gradients had the best IOU scores, but were computationally expensive. Grad-CAM and Occlusion could be used instead with slightly worse results. SHAP and LIME performed the worst, but this could be due to the generation of the bounding boxes from the segmented attribution maps. Future work could improve the bounding box generation process, and apply the XAI methods for breast cancer subtype classification.

REFERENCES

van der Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A., 2022. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. Medical Image Analysis 79, 102470. doi:https://doi.org/10.1016/j.media.2022.102470.

de Vries, B.M., Zwezerijnen, G.J.C., Burchell, G.L., van Velden, F.H.P., Menke-van der Houven van Oordt, C.W., Boellaard, R.,2023. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review. Front Med (Lausanne) 10, 1180773.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., Reblitz-Richardson, O., 2020. Captum: A unified and generic model interpretability library for pytorch.