



HAL
open science

Lesion detection in contrast enhanced spectral mammography

Clément Jailin, Pablo Milioni, Zhijin Li, Râzvan Iordache, Serge Muller

► **To cite this version:**

Clément Jailin, Pablo Milioni, Zhijin Li, Râzvan Iordache, Serge Muller. Lesion detection in contrast enhanced spectral mammography. Sixteenth International Workshop on Breast Imaging, May 2022, Leuven, Belgium. pp.24, 10.1117/12.2624577 . hal-03727549

HAL Id: hal-03727549

<https://hal.science/hal-03727549>

Submitted on 19 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lesion detection in Contrast Enhanced Spectral Mammography

Clément Jailin, Pablo Milioni, Zhijin Li, Răzvan Iordache, and Serge Muller

GE Healthcare, 78530 Buc, France

ABSTRACT

Background & purpose: The recent emergence of neural networks models for the analysis of breast images has been a breakthrough in computer aided diagnostic. This approach was not yet developed in Contrast Enhanced Spectral Mammography (CESM) where access to large databases is complex. This work proposes a deep-learning-based Computer Aided Diagnostic development for CESM recombined images able to detect lesions and classify cases.

Material & methods: A large CESM diagnostic dataset with biopsy-proven lesions was collected from various hospitals and different acquisition systems. The annotated data were split on a patient level for the training (55%), validation (15%) and test (30%) of a deep neural network with a state-of-the-art detection architecture. Free Receiver Operating Characteristic (FROC) was used to evaluate the model for the detection of 1) all lesions, 2) biopsied lesions and 3) malignant lesions. ROC curve was used to evaluate breast cancer classification. The metrics were finally compared to clinical results.

Results: For the evaluation of the malignant lesion detection, at high sensitivity ($Se > 0.95$), the false positive rate was at 0.61 per image. For the classification of malignant cases, the model reached an Area Under the Curve (AUC) in the range of clinical CESM diagnostic results.

Conclusion: This CAD is the first development of a lesion detection and classification model for CESM images. Trained on a large dataset, it has the potential to be used for helping the management of biopsy decision and for helping the radiologist detecting complex lesions that could modify the clinical treatment.

Keywords: Contrast enhanced spectral mammography, Breast cancer mass detection, Deep learning, Computer Aided Detection

1. INTRODUCTION

The recent development of Computer Aided Detection (CADe) and Computer Aided Diagnostic (CADx) based on Deep Learning (DL) has been a breakthrough in medical imaging¹⁻³ and in breast cancer analysis. In breast images, various CAD exist for screening or diagnostic applications and are used on Full Field Digital Mammography^{4,5} or Digital Breast Tomosynthesis⁶ where large databases may be available for the training.

Contrast-enhanced spectral mammography (CESM) provides anatomical and functional imaging of breast tissue improving the accuracy of breast cancer diagnosis.^{7,8} This recent imaging technique has started to attract broad research and clinical application interest.⁹⁻¹¹ After an intravenous iodine injection, the compressed breast is imaged with two X-ray energies. The iodine-equivalent image (called recombined image) is obtained from processing of the low and high energy images.

In Contrast Enhanced Digital Mammography application, the current developments in artificial intelligence are focused on lesion classification to predict the pathology results. Due to the small size of available databases (generally less than 130 patients), the approaches are often based on the classification of manually extracted regions of interests. From those images, handcrafted features (*e.g.*, radiomics¹²⁻¹⁵) or deep learning features

C.J.: E-mail: clement.jailin@ge.com

(*e.g.*, obtained from a convolutional neural network¹⁶) are extracted and fed into machine learning algorithms (*e.g.*, support vector machine, multi-layer perceptron, etc.) to perform the classification. Those analyses require the suspicious area to be detected and extracted manually by the user. In some applications,^{16,17} a pixel segmentation of the contrast uptakes may be necessary.

The current study is a first development of an automatic lesion detection and analysis model in CESM images. A deep learning model is trained on a large database (586 patients) of diagnostic CESM patients to localize contrast uptakes in the iodine images and associates them a suspicion score that would aid the clinician in their diagnostic. The results are finally compared with clinical diagnostic results.

2. DATA AND METHOD

CESM datasets A CESM dataset consisting of 586 patients with 2510 recombined images of left/right views, mainly cranio-caudal (CC) and mediolateral oblique (MLO), was collected from various hospitals and acquired with different systems (Senographe DS, Essential and Pristina from *GE Healthcare, Chicago, Illinois, United States*). The lesions in the dataset were all biopsy proven. The number of normal, benign and malignant cases are respectively 191 (33%), 149 (25%) and 246 (42%). When training a detection model, having normal cases allows reducing the false positive detection by training to recognize what is not a lesion. It is expected that a too large proportion of normal cases may also degrade the model efficiency. The optimal quantity of normal / annotated data was not studied in this paper. Note that for the cases where a lesion is detected, the contralateral breast may be labeled normal if no findings are detected. It therefore represents normal cases for the training. The summary of the data is provided in table 1.

Hospital	Num of patients	Acq. sys.	Pathology results		
			Normal	Benign	Malignant
Peking University First Hospital & Shanghai First People's Hospital (China)	244 (976 images)	DS / Essential	35 (14%)	64 (26%)	145 (60%)
CBIS - Carolina Breast Imaging Specialists (US)	26 (143 images)	Pristina	1 (5%)	5 (19%)	20 (77%)
Beth Israel Deaconess Medical Center (US)	50 (211 images)	Essential	2 (4%)	24 (48%)	24 (48%)
University of Cambridge (UK)	39 (156 images)	Pristina	0 (0%)	11 (28%)	28 (72%)
University of Washington (US)	187 (790 images)	Essential	153 (81%)	30 (2%)	4 (81%)
Hospital del Mar (Spain)	40 (234 images)	Pristina	0 (0%)	15 (38%)	25 (62%)

Table 1. Details on the CESM dataset and pathology results.

For the training of the deep learning models, the 586 patients data were split in a training (332 patients - 1501 images) / validation (82 patients - 332 images) / test sets (172 patients - 677 images) stratified by the pathology results (normal/benign/malignant) and by clinical sites. In addition to the biopsied lesions, all other benign lesions in the breast (*e.g.*, non-biopsied cysts, fibroadenoma, nodes, etc.) were annotated with a rectangular bounding box and labeled with a specific label.

All CESM recombined data were processed using the latest *GE Healthcare* commercial CESM recombination algorithm (Nira processing) to reduce artifacts. To be used in the deep learning framework, the images were pre-processed. The image intensities were bounded between 1950 and 2205, corresponding to most of the recombined breast and iodine intensity range and converted into 8 bits.

Method: The deep learning detection model used in this study is a small Yolo-v5s model. This is a PyTorch implementation of the similar Yolo-v4,¹⁸ a state-of-the-art model for object detection. With a scalable architec-

ture, it can be trained with relatively small datasets. From the published Yolo-v4, the only evolution of v5 used in this study is the automatic anchor size estimation based on k-means.

The architecture consists in 3 parts: (a) a CSPDarknet backbone to extract features at different scales, (b) a PANet neck to perform features combination using a feature pyramidal structure and (c) a Yolo Layer head that generates feature maps at three different scales allowing to identify small, intermediate and large lesions. The multi-scale characteristics of this model ensures a correct determination of all the lesion sizes (from a diameter < 3 mm up to > 10 cm).

We used as loss function $\mathcal{L}_{\text{DIoU}}$ a Distance Intersection over Union (DIoU) loss¹⁹ that simultaneously considers the overlapping area and the distance to the center of the ground truth boxes. The IoU is defined as the ratio between the area of overlap over the area of union between the detected and ground truth bounding boxes.

$$\mathcal{L}_{\text{DIoU}} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{x}, \mathbf{x}^{gt})}{c^2} \quad (1)$$

with ρ the Euclidean distance, \mathbf{x} and \mathbf{x}^{gt} respectively the coordinates of the centers for the predicted and ground truth boxes and c the diagonal length of a rectangle circumscribed to the predicted and ground truth boxes. This loss leads to a better convergence than simple IoU loss. It can be noted that other metrics (*e.g.*, IoU, Complete IoU, Generalized IoU) were tested yet not providing better results than DIoU. Complete IoU (CIoU), considering the aspect ratio of the bounding boxes, gave similar results certainly because the shapes of the annotations are not very discriminant in lesion detection applications.

To compensate for small (compared to Deep learning standards) datasets, two techniques were applied: data-augmentation strategies and transfer learning. Data-augmentation strategies suited to breast images were used such as image flips, global intensity transforms and dedicated breast geometrical realistic transforms as developed in a previous study.²⁰ The weights of the model were initially pre-trained on Image-Net, then re-trained on the lesion detection. Finally, the inference is performed with Test Time Augmentation (TTA) that consists in performing the inference on an image duplicated with simple transforms (scales and flips) and averaging all detections. After the inference, a non-maximum suppression (NMS) with a IoU threshold of 0.2 is used to reduce the number of superimposed detections.

The optimization was performed with a stochastic gradient descent ($lr = 0.01$, momentum= 0.937). The batch size was set to 12 images to fit the GPU memory (24Gb - Quadro RTX-6000). Approximately 500 iterations were necessary to reach convergence.

To validate the detection of an ROI, an acceptance criterion based on the IoU with the ground truth is used (threshold sets to $\text{IoU}_{\text{threshold}} = 0.3$). The results for the lesion detection are evaluated with FROC (Free Receiver Operating Characteristic) curves considering 3 metrics:

1. the sensitivity for the detection of all lesions (biopsied and not) in the breast
2. the sensitivity for the detection of only biopsied lesions (identified suspicious by the radiologist). A clinical use of this metric would be suspicious lesion detection in screening CESM. This application may help the radiologist for the analysis of complex cases (*e.g.*, small and blurry lesion, satellite lesion).
3. the sensitivity for the detection of cancers. This metric corresponds to a CAD aiming to reduce negative biopsies.

In addition, the predicted suspicion score associated to each detected lesion allows performing a more general classification at image, breast or patient level. Each breast suspicion score is hence defined by the score of the most suspicious detected lesion. This classification will be evaluated in term of sensitivity and specificity with a ROC curve and compared with results from radiologists extracted from a CESM diagnostic clinical review paper.²¹

3. RESULTS

The different FROC curves corresponding to the 3 defined metrics are plotted in figure 1(a). With a score threshold at 0.1 corresponding to a sensitivity of 0.95 for the detection of cancers (corresponding to the metric 3.), the false positive rate is 0.61 FP per breast. This metric is presented by the red FROC curve. The maximal detection sensitivity for cancers is 0.989 (with all detection scores accepted) with 3.8 FP per breast.

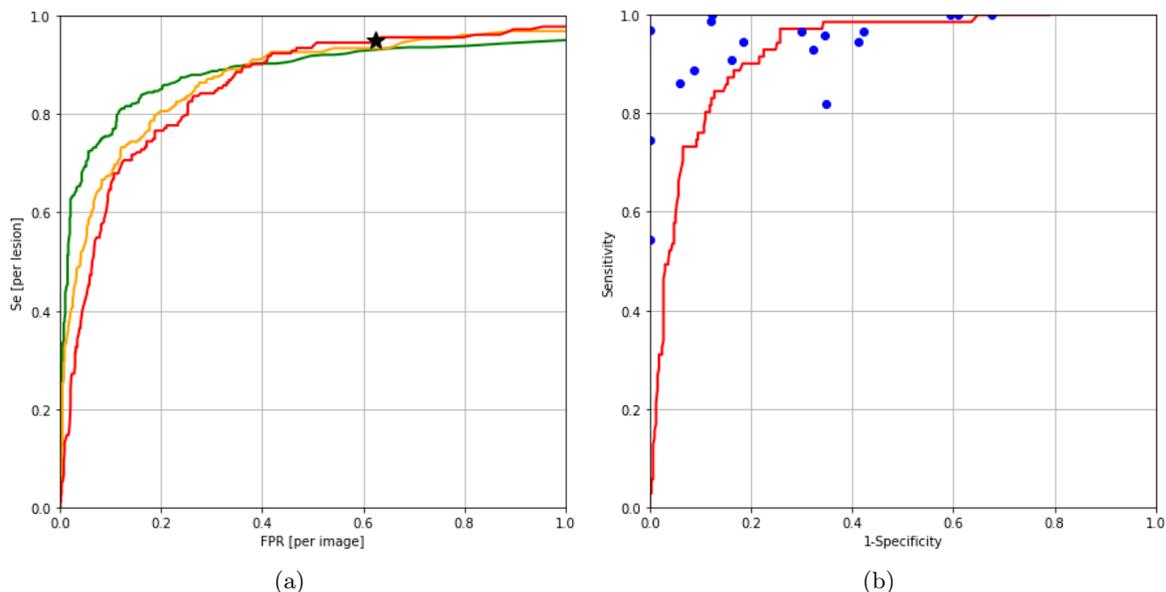


Figure 1. (a) FROC curve with the metrics defined in section 2 - green: 1, - orange: 2 and - red: 3. A marker is plotted for a sensitivity of 0.95 (star). (b) ROC curve for cancer classification at the breast level (blue dots correspond to radiologists' diagnostic results from²¹).

For the detection of cancers, the distribution of the model scores for true positive (TP) or false positive (FP) detections are shown figure 2 with respectively the red and blue bars. When a FP is detected, it is interesting to separate (black) the non-annotated detection, (green) the non-biopsied annotated lesions and (orange) the biopsied benign lesions. With a high detection score (>0.7), the FP are mainly composed of suspicious lesions that turned out to be benign after biopsy (orange bar). As they were labeled 'suspicious' the radiologist biopsied them and it was expected that those lesions could be confused with the malignant ones.

Figure 1(b) corresponds to the ROC curve for the cancer classification at the breast level and the blue dots are clinical results obtained from various radiologists classifying CESM images.²¹ The AUC of the deep learning breast classification model is 0.930. The median specificity for the 18 clinical studies is 0.757. At this fixed specificity, the sensibility of the AI model is 0.929.

The lesion detection results obtained for few patients are illustrated in figure 3 with the red boxes being the detected regions and the green boxes the annotated ground truth. It can be seen that the detections are mainly consistent in the CC and MLO views (similar structure detection when the lesion was visible in the two views). However, the model was designed to detect lesions for each view independently. In the last presented patient for example, a FP on the LMLO view was not seen on the LCC view. Developing a multi-view CAD would be of upmost interest to increase the CAD efficiency.

Figure 4 presents 3 views from three patients with FP detections. On those images the ground truth lesions are shown in green and the detected boxes in red. Although the ground truth lesions were correctly identified, multiple false positives are also detected. When analyzing the FP detections, three main categories appear:

- detection of localized non-lesion patterns (*e.g.*, vessels, biopsy clips, foreign objects, mole markers, etc.). An example of a non-lesion detection is shown figure 4 (left).

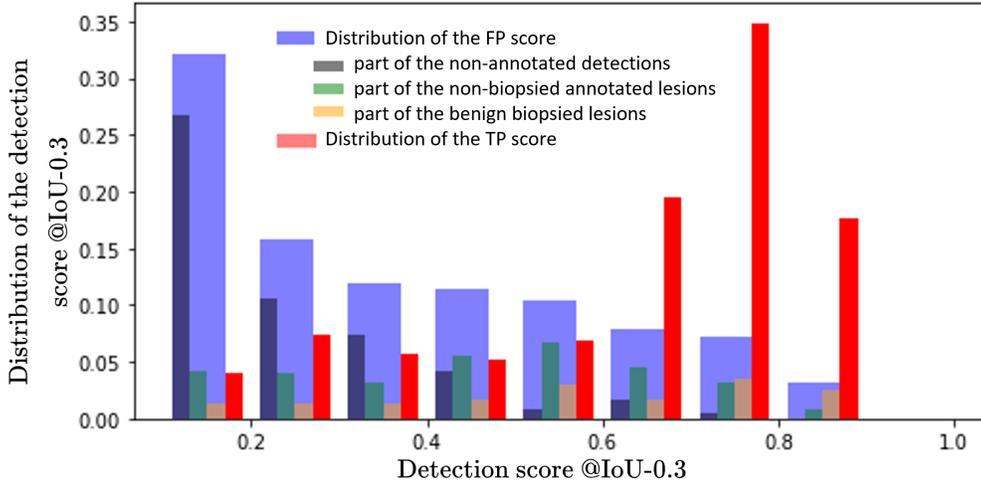


Figure 2. Distribution of the TP (red) and FP (blue) scores (in the range [0.1-1]) for all detections on the test set. The FP are separated into three categories: non-annotated lesions, annotated non-biopsied and annotated biopsied benign lesions. A score of 0.1 corresponds to 95% detection sensitivity.

- detection of fragmented or complex lesions where the annotation is inconsistent with the detection (*e.g.*, a micro-nodular contrast uptake annotated as a single box and detected with multiple ROIs). An example of this FP is shown figure 4 (center and right). In some cases, all fragmented lesions are detected and supplemented by a broad general annotation. These kinds of inconsistent detections may generate a lot of FP.
- detection of CESM artifacts (*e.g.*, scars, skin folds). For this reason, it is very important to use CESM recombined images with the least artifacts (*e.g.*, processed with Nira algorithm).

4. CONCLUSION

This study presents the development of a deep learning-based detection model in CESM images. A large CESM dataset composed of 586 patients (for a total of 2510 images) collected from various hospitals and acquired with different imaging systems from GE Healthcare is used. All labels were biopsy proven. The detection model is a small Yolo-v5 architecture, trained with data augmentation strategies. Various metrics were used to assess the efficiency of the detection and image classification. The trained model performed breast classification with an AUC of 0.93 and was able to identify cancers with a sensitivity of 0.95 and 0.61 false positive per image.

One perspective of this work is to include multi-view detection²² as our current model detects lesions independently of views. Unifying the detections should improve model performance.

Considering both the recombined and the low energy images will be a future investigation. As performed in CESM ROI classification papers,¹⁵ the low energy may contain discriminant information to classify the detected ROIs.

The model we developed has the potential to be used for triaging patients for biopsy or follow-up, and to help the radiologist detecting complex contrast uptake that could modify the treatment planning (*e.g.*, detection of satellite lesions). With already promising results, we aim to extend this study to a larger database to continue improving the model efficiency and robustness.

The originality of this study lies in two contributions:

- While deep learning models have been a breakthrough for lesion detection and analysis in FFDM and DBT images, they have not been employed in CESM. This work is, to the best of our knowledge, the first attempt to develop a deep-learning CAD for CESM.

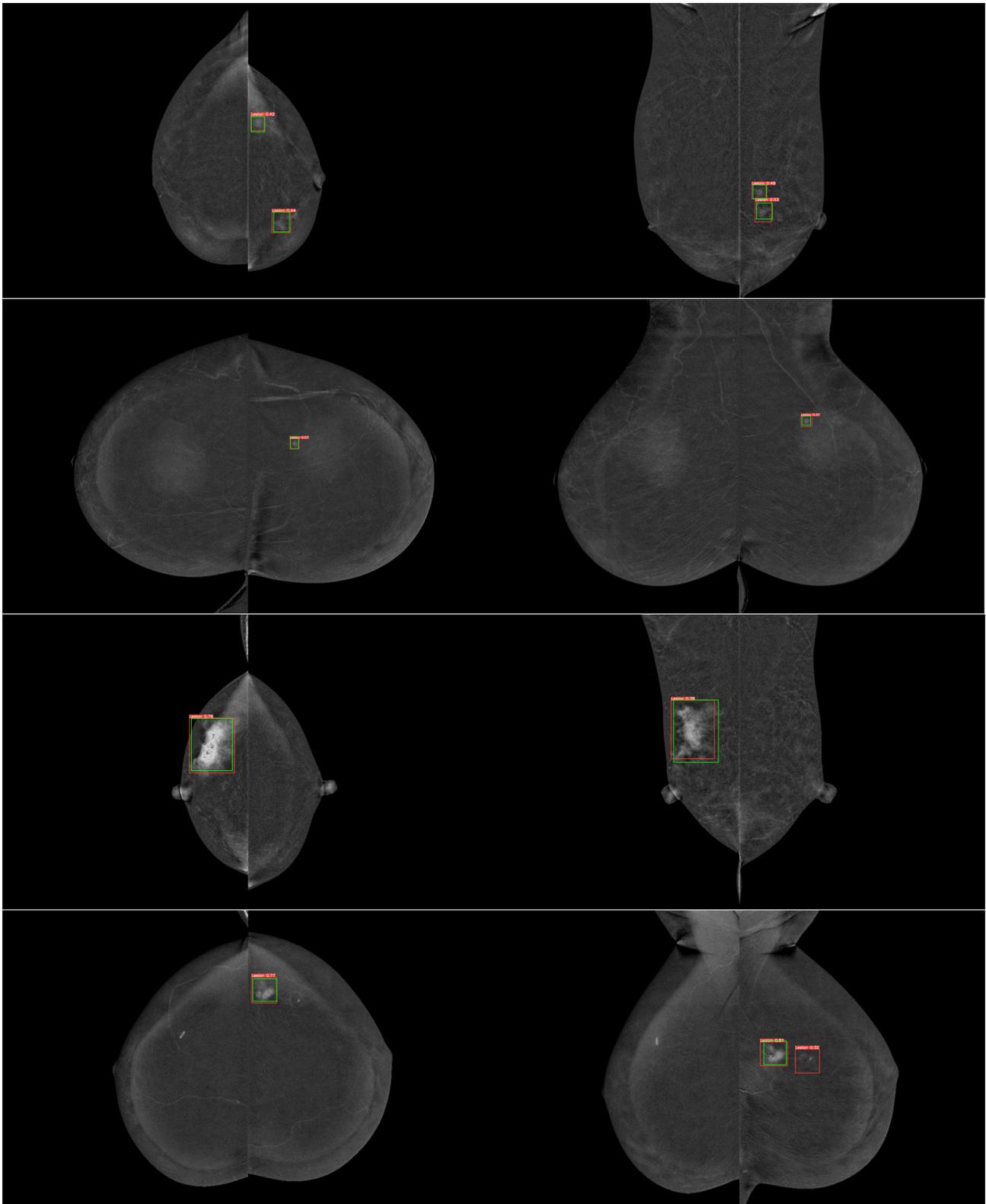


Figure 3. Results for lesion detection at $Se=0.95$. Red box: detected area, green box: ground truth. In addition, a suspicion score is associated to each ROI.

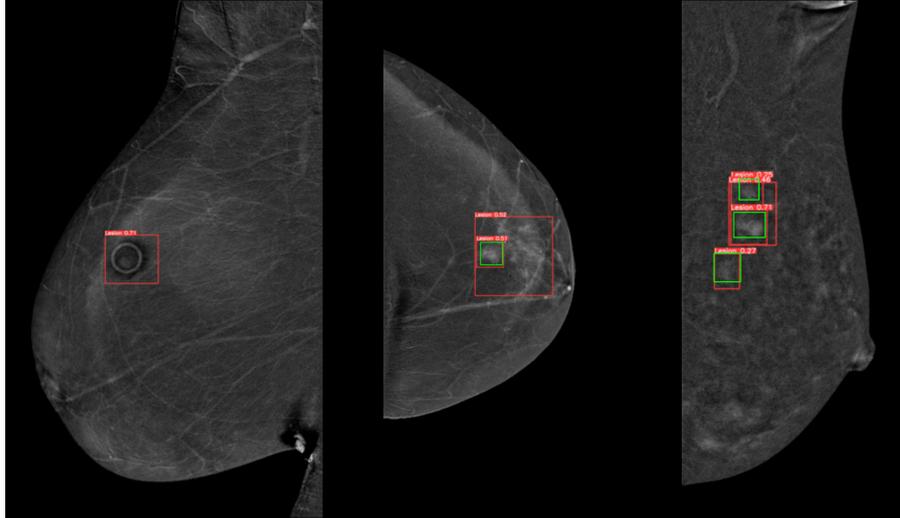


Figure 4. Three images containing FP detections (score threshold for cancer detection $Se=0.95$). (left) detection of a foreign object, (center) two detections of the contrast uptake, the smallest corresponds to the annotated ground truth while the largest is considered as a FP and (right) the two detected lesions are supplemented with another third large box that is considered as a FP.

- Collected from various sites, this CESM dataset with ground truth used in this study is remarkably large compared to the literature.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge Ruben Sanchez, Ann-Katherine Carton, Jean-Paul Antonini (all working at GE Healthcare, Buc, France) for their help in data collection and insights in mammography and Andrei Petrovskii for useful discussion on deep learning models. We wish to point out the absence of conflicts of interest related to our study.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using anonymized human subject data made available by research partners. Applicable law and standards of ethic have been respected.

REFERENCES

- [1] S. Zhou, H. Greenspan, D. Shen, Deep learning for medical image analysis, Academic Press, 2017.
- [2] B. Sahiner, A. Pezeshk, L. Hadjiiski, X. Wang, K. Drukker, K. Cha, R. Summers, M. Giger, Deep learning in medical imaging and radiation therapy, *Medical physics* 46 (1) (2019) e1–e36.
- [3] H. Fujita, AI-based computer-aided diagnosis (AI-CAD): the latest review to read first, *Radiological physics and technology* 13 (1) (2020) 6–19.
- [4] L. Shen, L. Margolies, J. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, *Scientific reports* 9 (1) (2019) 1–12.
- [5] H.-E. Kim, H. H. Kim, B.-K. Han, K. H. Kim, K. Han, H. Nam, E. H. Lee, E.-K. Kim, Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study, *The Lancet Digital Health* 2 (3) (2020) e138–e148.
- [6] M. A. Al-Masni, M. A. Al-Antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, T.-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, *Computer methods and programs in biomedicine* 157 (2018) 85–94.

- [7] C. Dromain, F. Thibault, S. Muller, F. Rimareix, S. Delalogue, A. Tardivon, C. Balleyguier, Dual-energy contrast-enhanced digital mammography: initial clinical results, *European radiology* 21 (3) (2011) 565–574.
- [8] E. Luczyńska, S. Heinze-Paluchowska, S. Dyczek, P. Blecharz, J. Rys, M. Reinfuss, Contrast-enhanced spectral mammography: comparison with conventional mammography and histopathology in 152 women, *Korean journal of radiology* 15 (6) (2014) 689–696.
- [9] F. Diekmann, S. Diekmann, F. Jeunehomme, S. Muller, B. Hamm, U. Bick, Digital mammography using iodine-based contrast media: initial clinical experience with dynamic contrast medium enhancement, *Investigative radiology* 40 (7) (2005) 397–404.
- [10] S. Puong, X. Bouchevreau, F. Patoureaux, R. Iordache, S. Muller, Dual-energy contrast enhanced digital mammography using a new approach for breast tissue canceling, in: *Medical Imaging 2007: Physics of Medical Imaging*, Vol. 6510, International Society for Optics and Photonics, 2007, p. 65102H.
- [11] J. M. Lewin, B. K. Patel, A. Tanna, Contrast-enhanced mammography: a scientific review, *Journal of Breast Imaging* 2 (1) (2020) 7–15.
- [12] L. Losurdo, A. Fanizzi, T. M. A. Basile, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, V. Lorusso, R. Massafra, P. Tamborra, et al., Radiomics analysis on contrast-enhanced spectral mammography images for breast cancer diagnosis: A pilot study, *Entropy* 21 (11) (2019) 1110.
- [13] B. K. Patel, S. Ranjbar, T. Wu, B. A. Pockaj, J. Li, N. Zhang, M. Lobbes, B. Zhang, J. R. Mitchell, Computer-aided diagnosis of contrast-enhanced spectral mammography: A feasibility study, *European journal of radiology* 98 (2018) 207–213.
- [14] F. Lin, Z. Wang, K. Zhang, P. Yang, H. Ma, Y. Shi, M. Liu, Q. Wang, J. Cui, N. Mao, et al., Contrast-enhanced spectral mammography-based radiomics nomogram for identifying benign and malignant breast lesions of sub-1 cm, *Frontiers in Oncology* 10 (2020) 2407.
- [15] R. Massafra, S. Bove, V. Lorusso, A. Biafora, M. C. Comes, V. Didonna, S. Diotaiuti, A. Fanizzi, A. Nardone, A. Nolasco, et al., Radiomic feature reduction approach to predict breast cancer by contrast-enhanced spectral mammography images, *Diagnostics* 11 (4) (2021) 684.
- [16] S. Perek, N. Kiryati, G. Zimmerman-Moreno, M. Sklair-Levy, E. Konen, A. Mayer, Classification of contrast-enhanced spectral mammography (CESM) images, *International journal of computer assisted radiology and surgery* 14 (2) (2019) 249–257.
- [17] M. Caballo, D. R. Pangallo, W. Sanderink, A. M. Hernandez, S. H. Lyu, F. Molinari, J. M. Boone, R. M. Mann, I. Sechopoulos, Multi-marker quantitative radiomics for mass characterization in dedicated breast CT imaging, *Medical physics* 48 (1) (2021) 313–328.
- [18] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934 (2020).
- [19] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, 2020, pp. 12993–13000.
- [20] L. Caselles, C. Jailin, S. Muller, Data augmentation for breast cancer mass segmentation, in: *International Conference on Medical Imaging and Computer-Aided Diagnosis*, Springer, 2021, pp. 228–237.
- [21] X. Zhu, J.-M. Huang, K. Zhang, L.-J. Xia, L. Feng, P. Yang, M.-Y. Zhang, W. Xiao, H.-X. Lin, Y.-H. Yu, Diagnostic value of contrast-enhanced spectral mammography for screening breast cancer: systematic review and meta-analysis, *Clinical breast cancer* 18 (5) (2018) e985–e995.
- [22] Z. Yang, Z. Cao, Y. Zhang, Y. Tang, X. Lin, R. Ouyang, M. Wu, M. Han, J. Xiao, L. Huang, et al., MommiNet-v2: Mammographic multi-view mass identification networks, *Medical Image Analysis* 73 (2021) 102204.