



Intra-class Contrastive Learning Improves Computer Aided Diagnosis of Breast Cancer in Mammography

Kihyun You¹ , Suho Lee^{1,2} , Kyuhee Jo^{1,3} , Eunkyung Park¹, Thijs Kooi¹, and Hyeonseob Nam¹

¹ Lunit Inc., Seoul, Republic of Korea

{ukihyun,shlee,kjo,ekpark,tkooi,hsnam}@lunit.io

² Department of Data Science, Seoul National University of Science and Technology, Seoul, Republic of Korea

swlee@ds.seoultech.ac.kr

³ Johns Hopkins University, Baltimore, MD, USA

kjo3@jh.edu

Abstract. Radiologists consider fine-grained characteristics of mammograms as well as patient-specific information before making the final diagnosis. Recent literature suggests that a similar strategy works for Computer Aided Diagnosis (CAD) models; multi-task learning with radiological and patient features as auxiliary classification tasks improves the model performance in breast cancer detection. Unfortunately, the additional labels that these learning paradigms require, such as patient age, breast density, and lesion type, are often unavailable due to privacy restrictions and annotation costs. In this paper, we introduce a contrastive learning framework comprising a Lesion Contrastive Loss (LCL) and a Normal Contrastive Loss (NCL), which jointly encourage models to learn subtle variations beyond class labels in a self-supervised manner. The proposed loss functions effectively utilize the multi-view property of mammograms to sample contrastive image pairs. Unlike previous multi-task learning approaches, our method improves cancer detection performance without additional annotations. Experimental results further demonstrate that the proposed losses produce discriminative intra-class features and reduce false positive rates in challenging cases.

Keywords: Mammography · Multi-task learning · Contrastive learning

1 Introduction

Mammography is the most common and cost-effective method for early detection of breast cancer—the second most common cancer in women worldwide [22].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16437-8_6.

Computer aided diagnosis (CAD) systems have been used as a way to assist radiologists in reading millions of mammograms [11]. Thanks to the abundance of screening data and the recent advance in deep neural networks, CAD applications for mammography have been developed rapidly; some papers demonstrate performance comparable to its human counterpart [19, 21].

Literature suggests that imitating how a radiologist reads a mammogram is a promising way to improve the performance of CAD models. For example, models that use multiple views (left and right or mediolateral oblique (MLO) and craniocaudal (CC)) in a mammographic exam outperform models that only work on a single view [12, 15, 17, 25]. Furthermore, multi-task learning, where a model is trained to predict various radiological features such as the Breast Imaging-Reporting and Data System (BI-RADS) category [23], breast density, radiological subtype of a lesion as well as patient features like age, has shown to be effective in improving the performance of cancer detection [12, 25]. However, this extra information is often unavailable due to privacy issues and huge annotation costs.

Contrastive learning is a training technique that can bypass this problem by learning useful features from the data themselves. Its goal is to pull similar examples (i.e., positive pairs) closer and push dissimilar examples (i.e., negative pairs) farther in the embedding space. Contrastive learning has recently surged in popularity for its success in self-supervised learning [4, 8] and supervised representation learning [7, 18, 20]. In medical image analysis, contrastive learning has been used to pre-train a triage system for mammograms [2] and to improve domain generalization across multiple vendors [14]. Yan et al. [24] uses contrastive learning to match lesion patches generated from two different views of a patient as a part of multi-task learning for mammography. Chen et al. [3] exploits anatomical symmetry of body parts to detect fraction on X-ray by pixel-level contrastive learning with flipped images.

In this paper, we propose two contrastive losses, a Lesion Contrastive Loss (LCL) and a Normal Contrastive Loss (NCL), that leverage the multi-view nature of mammograms to explore rich radiological features beyond binary (cancer vs. non-cancer) class labels. The LCL attracts embeddings of different views of the same lesion and repels embeddings of different lesions. Similarly, the NCL attracts embeddings of bilateral normal breasts and repels embeddings of normal breasts from different patients. By contrasting samples that have the same class but are from distinct patients, our method allows the model to indirectly learn variations within the class, i.e., the intra-class variations. These losses are optimized jointly along with the image-level cancer classification loss via multi-task learning.

The main contributions of this paper are summarized as follows. 1) We introduce LCL and NCL that enable the model to learn auxiliary information without explicit annotations. To the best of our knowledge, our framework is the first approach in mammography that exploits contrastive learning to model intra-class variations in a self-supervised manner. 2) We show that our losses improve the overall cancer detection performance of vanilla baseline model as well as previous multi-task learning approaches, demonstrating the general applicability and

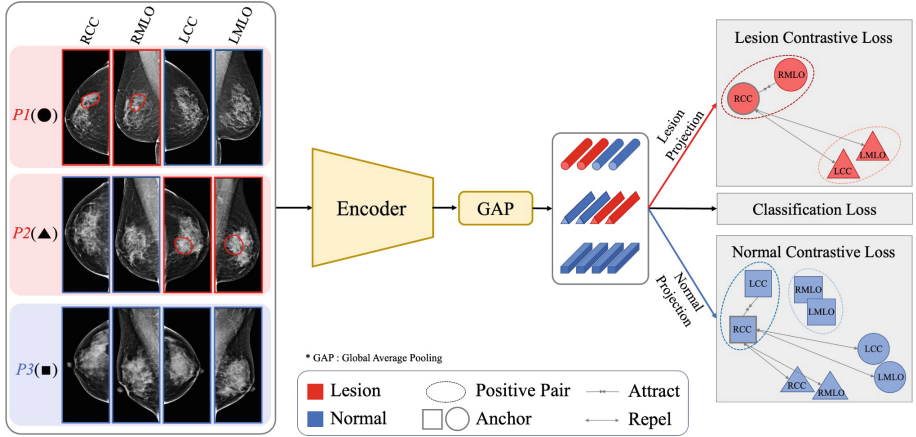


Fig. 1. Our contrastive learning framework consists of a Lesion Contrastive Loss (LCL) and a Normal Contrastive Loss (NCL). The LCL learns the similarity of a lesion seen at different views of a patient and the dissimilarity of lesions from different patients. The NCL, in contrast, utilizes parenchymal symmetry between both breasts of a normal patient and diversity of normal breasts from different patients.

efficacy of our method. 3) Extensive experiments on a large scale in-house dataset show that our method reduces false positive rates on challenging negative cases and produces features that are transferable to more fine-grained classification tasks. These results suggest that LCL and NCL combined generates features that more aptly reflect intra-class variations.

2 Method

Our work aims to exploit the four-view nature of mammograms (RCC, LCC, RMLO, and LMLO) to learn intra-class variations without using extra labels. Inspired by contrastive clustering [13], we introduce two contrastive losses—Lesion Contrastive Loss (LCL) and Normal Contrastive Loss (NCL)—that together encourage the model to explore intricate variability within class beyond the simple binary classification defined by the presence or absence of lesions. The overview of the proposed method is illustrated in Fig. 1.

2.1 Multi-task Learning

Following the success of multi-task learning in mammograms [12, 24], our method is built upon a multi-task learning framework comprising three losses: an image-level classification loss, LCL and NCL. As shown in Fig. 1, an input mammogram is first fed into a fully-convolutional encoder, which consists of a pre-trained convolutional neural network followed by additional convolutional layers to further abstract the output feature maps. Global average pooling (GAP) is subsequently applied to reduce the feature maps into a 1-dimensional vector. The produced

feature vector is then passed to a classification head for distinguishing cancer from non-cancer, and two projection heads to produce embedding vectors that are used for contrastive learning as described below.

2.2 Contrastive Losses

Contrastive loss is a function designed to learn an embedding space in which inputs with similar properties lie close to each other while inputs with distinct characteristics are placed distantly. One way to implement the contrastive loss is utilizing a unit of three images—a triplet—consisting of a reference image called an anchor (a), a matching image called a positive (p), and a non-matching image called a negative (n). We adopt the triplet loss proposed in [20]. Given triplets $(a, p, n) \in T$, where T is the entire set of triplets, the triplet loss is formulated as

$$L_{\text{triplet}} = \frac{1}{|T|} \sum_{(a,p,n) \in T} \max(d(a,p)^2 - d(a,n)^2 + m, 0), \quad (1)$$

where m is a margin and $d(i, j)$ is the L2 distance between two images i and j on the embedding space. The triplet loss tries to keep the distance from the anchor (a) to the positive (p) smaller than the distance to the negative (n) by a margin m . We apply this formula to both LCL and NCL.

Lesion Contrastive Loss. Radiologists read mammograms not only by assessing the malignancy of a lesion (e.g. BI-RADS category) but also by describing various radiological features of the lesion such as the lesion type (e.g. mass, calcification, asymmetry, and distortion), shape (e.g. oval, round, and irregular), size, density, etc. These features are not captured when a model is trained merely with malignancy labels such as BI-RADS categories or biopsy results.

Motivated by this, we introduce the Lesion Contrastive Loss (LCL) that indirectly learns the various aspects of lesions by comparing breasts containing lesions in a self-supervised manner. In mammogram screening, lesions often in two different images called CC and MLO views. In LCL, the two images containing the same lesion (e.g. RCC and RMLO from a patient having cancer in the right breast) are pulled together, while images of different lesions (e.g. RCCs of two patients having cancer in their right breasts) are repelled from each other. Let an image containing a lesion be an anchor a . The ipsilateral image of the same lesion acts as a positive (p) and other patient’s image that also contain a lesion (hence a different lesion) act as a negative (n) in Eq. 1

Normal Contrastive Loss. Normal breasts (i.e., breasts that contain neither a malignant nor benign lesion) exhibit radiological variations such as parenchymal patterns (breast density), unsuspicious benign patterns, surgical scars, and other textural characteristics. Literature indicates that these features may associate with the risk of cancer [1]. Additionally, a normal contralateral breast of a positive exam has a higher risk of developing cancer [10, 26].

The normal Contrastive Loss (NCL) is designed to make better use of this latent variability among normal breasts. Based on the assumption that bilateral breasts are anatomically symmetric [3], the NCL draws together the images of left and right breasts of a normal patient (e.g. RCC and LCC of a normal patient) and pushes apart normal images from different patients (e.g. RCCs of two normal patients). Let an image without a lesion, i.e., a normal sample, to be an anchor (a). The corresponding view of its opposite breast acts as a positive (p), and an image of normal breast from a different patient acts as a negative (n).

2.3 Hard Negative Sampling

The proposed triplet strategy requires all four views of each case to be present in a single mini-batch. To this end, we draw batches on a case level. Furthermore, to ensure unbiased training, we maintain class balance by sampling an equal number of lesion (cases including a cancer or benign lesion) and non-lesion cases (cases not including a lesion) for each batch.

Naive sampling of triplets in a mini batch can be problematic because the negative pairs are often trivial. In other words, the negative sample is too distant from the anchor to enter the margin of the triplet loss and yield a significant training signal. To address this problem, we employ Hard Negative Mining (HNM) to select effective triplets that violate the margin constraint. Given an anchor image, we calculate pairwise L2 distances to all candidate negative images in the mini batch and select the five closest images as negatives. This not only accelerates training but also improves the final performance as demonstrated in ablation studies.

3 Experiments

3.1 Dataset and Metrics

We construct an in-house mammography dataset consisting of 88,753 exams of four-view full-field digital mammograms from 10 institutions in two countries; eight in South Korea and two in the USA. Among the exams, 11,276 are cancer positive confirmed by biopsy, 36,636 are benign confirmed by biopsy or at least one year of follow-up, and 40,841 are normal confirmed by at least one year of follow-up. The mammograms are taken using devices from three manufacturers: GE, Hologic and Siemens with ratio 48.9%, 46.1%, and 5.0% respectively. We sample a validation set of 2,943 (986 cancer, 963 benign, 994 normal) exams and a test set of 2,928 (953 cancer, 991 benign, and 984 normal)¹ exams to evaluate models.

We compute three metrics to evaluate the results: area under the receiver operating characteristic curve (AUROC), sensitivity at a fixed specificity (0.8), and specificity at a fixed sensitivity (0.8). The DeLong test [5] is used to generate confidence intervals for AUROC and a p-value when comparing two methods (Fig. 2).

¹ We randomly sampled 1,000 exams per category for each validation and test set, and a few outlier exams (e.g. breast implants) are excluded.

Table 1. Cancer diagnosis performance (AUROC with 95% confidence interval and p-value, sensitivity at specificity 0.8 and vice versa) of our method compared with three settings. Our method improves every baseline with a statistically significant margin.

Method	AUROC	p-value	Sens.	Spec.
Baseline	0.862 ± 0.010	-	0.760	0.742
+Ours	0.878 ± 0.009	<0.0001	0.781	0.773
Multi-task learning of [25]	0.867 ± 0.010	-	0.764	0.750
+Ours	0.880 ± 0.009	<0.0001	0.780	0.775
Multi-task learning of [12]	0.876 ± 0.009	-	0.782	0.778
+Ours	0.885 ± 0.009	<0.0001	0.799	0.798

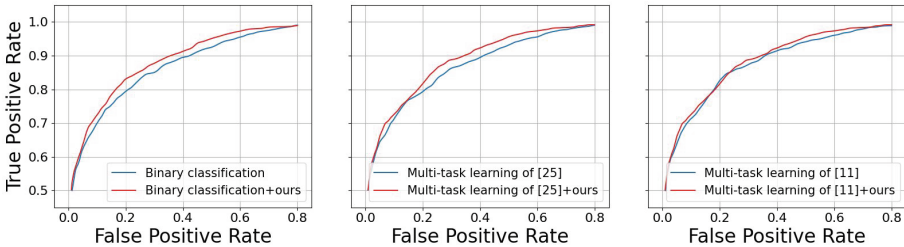


Fig. 2. ROC curves of each baseline and our method.

3.2 Implementation Detail

Our method is implemented in PyTorch, and trained with four GPUs of NVIDIA V100. We adopt ResNet-34 [9] pre-trained on ImageNet [6] as the backbone and use two additional convolution layers followed by global average pooling. A sigmoid classification layer is applied to predict a cancer score which is used to calculate binary cross-entropy loss. The projection layer for each of our contrastive losses is composed of 1×1 convolution and L2 normalization. We trained the models for 200K iterations with SGD where the learning rate, weight decay, and momentum are set to 0.005, 0.0001, and 0.9, respectively; the learning rate follows the cosine annealing learning rate scheduling [16].

During training, the input images are resized to 960×640 and the batch size is set to 128 (32 exams and 4 views per exam, as mentioned in Sect. 2.3) to fit our GPU memory constraint. To avoid overfitting, we augment images with various geometric transformations (translation, scaling, rotation, etc.) and photometric transformations (brightness, contrast, noise, etc.). For two contrastive losses, we use the triplet loss [20] with weight coefficient 5 and margin 0.2. We also investigate two more forms of contrastive loss: pair-wise contrastive loss [7] and InfoNCE loss [18]. Hyperparameters for these losses are specified in the ablation study in Sect. 3.3.

Table 2. Ablation analysis on the type of contrastive loss and the three components of our method. HNM means hard negative mining in Sect. 2.3.

Type of loss	LCL	NCL	HNM	AUROC	p-value	Sens.	Spec.
None (Baseline)				0.862 ± 0.010	-	0.760	0.742
Triplet [20]	✓			0.871 ± 0.010	0.0007	0.763	0.751
		✓		0.869 ± 0.010	0.0061	0.768	0.759
	✓	✓		0.874 ± 0.010	<0.0001	0.778	0.769
	✓	✓	✓	0.878 ± 0.010	<0.0001	0.781	0.773
Pairwise loss [7]	✓	✓	✓	0.873 ± 0.010	<0.0001	0.776	0.767
InfoNCE [18]	✓	✓	✓	0.876 ± 0.009	<0.0001	0.777	0.768

3.3 Result and Analysis

Performance Comparison. We evaluate the proposed loss functions by adding them to three settings: 1) a naive baseline with binary classification between cancer and non-cancer (i.e., benign and normal), 2) the multi-task learning setting proposed in MommiNet-v2 [25] which trains with BI-RADS categories as an additional regression task, and 3) the multi-task learning framework proposed in MVMT [12] which utilizes abundant extra information such as the BI-RADS category, breast density, radiological subtype (i.e., mass, micro-calcification, etc.), conspicuity, and patient age for additional tasks.

As shown in Table 1, adding auxiliary tasks using additional labels tends to improve the performance of the models. Notably, our unsupervised, label-free intra-class learning method outperforms supervised multi-task learning approaches in terms of the AUROC score. Furthermore, our method boosts the performance of previous multi-task approaches even further, which implies that LCL and NCL encourage models to learn orthogonal characteristics that are not captured in supervised multi-task learning methods.

Ablation Study. We perform an ablation analysis to demonstrate the contribution of each component (LCL, NCL, and HNM) and the choice of the specific implementation of contrastive loss. Table 2 shows that each part of the proposed framework contributes to the increased performance. Furthermore, our method achieves improvement regardless of the specific type of contrastive loss: triplet loss [20], pairwise loss [7] (weight 10 and margin 0.2), and InfoNCE [18] (weight 0.2 and temperature coefficient 2). The triplet loss yields the best performance.

Feature Analysis. We analyze the feature representation trained with our method by quantifying its capability to learn intra-class distinction. We train supervised linear classifiers on top of the trained features for two downstream tasks, lesion type (mass/asymmetry/distortion/micro-calcification) and breast density (A/B/C/D) predictions, and measure accuracy on the test set. Labels for these analyses were provided by board-certified radiologists on the in-house dataset. As shown in Table 3, our method significantly outperforms the baseline in both tasks with very low p-values, under 10^{-5} , which demonstrates that

Table 3. Linear classification accuracy (%) with trained feature representations. Results are shown with 95% confidence interval.

Method	Lesion type	Breast density
Baseline	53.9 ± 0.053	75.1 ± 0.030
+ Ours	58.2 ± 0.055	78.1 ± 0.029

Table 4. Subgroup analysis in terms of false positive rate (%) at sensitivity 0.8 with three different negative groups. Our method effectively reduces false positives in difficult negative groups.

Negatives	All	Biopsy-proven benign	Contralateral breast
Baseline	20.70	47.51	42.89
+ Ours	16.41	40.33	35.51

the learned representations effectively reflect intra-class variations among lesion breasts and normal breasts.

Subgroup Analysis. We perform a subgroup analysis on groups of data that have previously been prone to false positive predictions: (1) a set of biopsy-proven benign lesions and (2) a set of contralateral breasts (the opposite side to the breast diagnosed with biopsy-proven malignancy). The size of each negative subset is 181 and 949 cases respectively. A biopsy-proven benign lesion is not cancerous but is suspicious enough to be recalled for a pathological examination and can therefore be seen as a challenging negative sample. Furthermore, a breast that is contralateral to one that contains cancer does not only share anatomical similarity with its malignant counterpart such as texture and breast density, but also has a higher risk of developing cancer [10, 26]. This, combined with the batch dependency in exam-based sampling, may undesirably cause models to associate contralateral breasts with the presence of cancer.

We show that the proposed loss functions mitigate false positive predictions for both subgroups. As shown in Table 4, our framework significantly reduces the false positive rate, demonstrating that learning intra-class variation helps handling ambiguous examples near the class boundary.

4 Conclusion

We presented an intra-class contrastive learning framework that learns auxiliary information without supervision. We proved the efficacy of learning an embedding space that reflects intra-class variations through extensive experiments with a total of 88,753 exams. Compared to supervised multi-task learning approaches, our LCL+NCL substantially improves the model performance without requiring additional labels. Furthermore, we showed that our losses help models to learn features that are transferable to more fine-grained classification tasks such

as breast density and lesion subtype prediction. Lastly, our proposed method reduces false positive rate on radiologically challenging cases (biopsy-proved benign) and contextually challenging cases (contralateral breast of a cancerous breast).

References

1. Boyd, N.F., et al.: Mammographic density and the risk and detection of breast cancer. *New Engl. J. Med.* **356**(3), 227–236 (2007). <https://doi.org/10.1056/NEJMoa062790>, pMID: 17229950
2. Cao, Z., et al.: Supervised contrastive pre-training for mammographic triage screening models. In: de Bruijne, M., et al. (eds.) *MICCAI 2021. LNCS*, vol. 12907, pp. 129–139. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_13
3. Chen, H., et al.: Anatomy-aware Siamese network: exploiting semantic asymmetry for accurate pelvic fracture detection in X-ray images. *CoRR* abs/2007.01464 (2020). <https://arxiv.org/abs/2007.01464>
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations (2020)
5. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 837–845 (1988)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
7. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), vol. 2, pp. 1735–1742 (2006). <https://doi.org/10.1109/CVPR.2006.100>
8. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.B.: Momentum contrast for unsupervised visual representation learning. *CoRR* abs/1911.05722 (2019). <http://arxiv.org/abs/1911.05722>
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *CoRR* abs/1512.03385 (2015). <http://arxiv.org/abs/1512.03385>
10. Hungness, E.S., et al.: Bilateral synchronous breast cancer: mode of detection and comparison of histologic features between the 2 breasts. *Surgery* **128**(4), 702–707 (2000)
11. Kim, H.E., et al.: Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit. Health* **2**(3), e138–e148 (2020). [https://doi.org/10.1016/S2589-7500\(20\)30003-0](https://doi.org/10.1016/S2589-7500(20)30003-0). <https://www.sciencedirect.com/science/article/pii/S2589750020300030>
12. Kyono, T., Gilbert, F.J., van der Schaar, M.: Multi-view multi-task learning for improving autonomous mammogram diagnosis. In: Doshi-Velez, F., et al. (eds.) *Proceedings of the 4th Machine Learning for Healthcare Conference. Proceedings of Machine Learning Research*, vol. 106, pp. 571–591. PMLR (2019). <https://proceedings.mlr.press/v106/kyono19a.html>
13. Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J.T., Peng, X.: Contrastive clustering. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, pp. 8547–8555 (2021). <https://ojs.aaai.org/index.php/AAAI/article/view/17037>

14. Li, Z., et al.: Domain generalization for mammography detection via multi-style and multi-view contrastive learning (2021)
15. Liu, Y., Zhang, F., Chen, C., Wang, S., Wang, Y., Yu, Y.: Act like a radiologist: towards reliable multi-view correspondence reasoning for mammogram mass detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 1 (2021). <https://doi.org/10.1109/TPAMI.2021.3085783>
16. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with restarts. *CoRR abs/1608.03983* (2016). <http://arxiv.org/abs/1608.03983>
17. Ma, J., Li, X., Li, H., Wang, R., Menze, B., Zheng, W.S.: Cross-view relation networks for mammogram mass detection. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 8632–8638 (2021). <https://doi.org/10.1109/ICPR48806.2021.9413132>
18. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. *CoRR abs/1807.03748* (2018). <http://arxiv.org/abs/1807.03748>
19. Salim, M., et al.: External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol.* **6**(10), 1581–1588 (2020)
20. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. *CoRR abs/1503.03832* (2015). <http://arxiv.org/abs/1503.03832>
21. Sechopoulos, I., Teuwen, J., Mann, R.: Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: state of the art. In: *Seminars in Cancer Biology*, vol. 72, pp. 214–225. Elsevier (2021)
22. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. *CA: Cancer J. Clin.* **72**(1), 7–33 (2022). <https://doi.org/10.3322/caac.21708>. <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21708>
23. Spak, D., Plaxco, J., Santiago, L., Dryden, M., Dogan, B.: BI-RADS® fifth edition: a summary of changes. *Diagn. Int. Imaging* **98**(3), 179–190 (2017). <https://doi.org/10.1016/j.diii.2017.01.001>. <https://www.sciencedirect.com/science/article/pii/S2211568417300013>
24. Yan, Y., Conze, P.H., Lamard, M., Quéllec, G., Cochener, B., Coatrieux, G.: Multi-tasking Siamese networks for breast mass detection using dual-view mammogram matching (2020). https://doi.org/10.1007/978-3-030-59861-7_32
25. Yang, Z., et al.: MommiNet-v2: mammographic multi-view mass identification networks. *Med. Image Anal.* **73**, 102204 (2021). <https://doi.org/10.1016/j.media.2021.102204>. <https://www.sciencedirect.com/science/article/pii/S1361841521002498>
26. Yi, M., et al.: Predictors of contralateral breast cancer in patients with unilateral breast cancer undergoing contralateral prophylactic mastectomy. *Cancer* **115**(5), 962–971 (2009)