

Supervised Contrastive Pre-Training for Mammographic Triage Screening Models

Zhenjie Cao¹, Zhicheng Yang¹, Yuxing Tang¹, Yanbo Zhang¹, Mei Han¹, Jing Xiao², Jie Ma³, and Peng Chang¹

¹ PingAn Tech, US Research Lab, Palo Alto, USA

² Ping An Technology, Shenzhen, China

³ Shenzhen People’s Hospital, Shenzhen, China
pengchang@gmail.com

Abstract. Inspired by the recent success of self-supervised contrastive pre-training on ImageNet, this paper presents a novel framework of Supervised Contrastive Pre-training (SCP) followed by Supervised Fine-tuning (SF) to improve mammographic triage screening models. Our experiments on a large-scale dataset show that the SCP step can effectively learn a better embedding and subsequently improve the final model performance in comparison with the direct supervised training approach. Superior results of AUC and specificity/sensitivity have been achieved for our mammographic screening task compared to previously reported SOTA approaches.

Keywords: Mammogram · Contrastive learning · Screening · Dual-view

1 Introduction

Mammographic screening is a cost-effective method for early detection of breast cancer, with approximately 39 million mammograms performed annually in the United States [1]. It has been reported that the U.S. radiologists ranged from 66.7% to 98.6% for sensitivity and from 71.2% to 96.9% for specificity in mammogram-based breast cancer diagnosis [19]. While many previous works proposed deep learning (DL) models to identify cancer patients and help improve the radiologists’ performance [21], in this paper, we focus on training DL models to triage a portion of mammograms as cancer-free to reduce radiologists’ workload, and therefore improve their efficiency and specificity, without sacrificing sensitivity.

Unlike the previous deep neural nets (DNN)-based mammographic screening systems, which are trained directly by supervised learning [4, 16, 20, 21, 25], we propose a Supervised Contrastive Pre-training + Supervised Fine-tuning (SCP+SF) framework. It first performs the SCP pre-training through a carefully designed Siamese contrastive learning module, searching for an ideal embedding space, then transfers the pre-trained encoder to the SF module for the supervised fine-tuning phase.

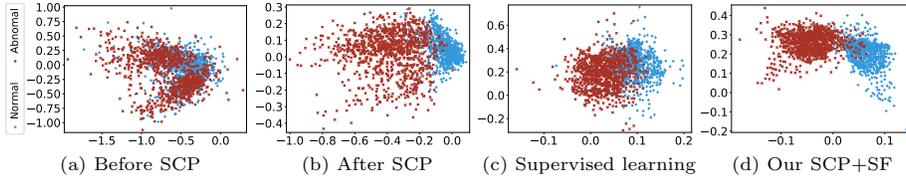


Fig. 1. Visualization of the normal and abnormal sample projections. (a)(b) are from the Siamese contrastive learning module, and (c)(d) are from the final dual-view model.

Contrastive learning has been applied to self-supervised visual representation learning [6, 7, 10, 11, 29], exemplified by the recent success of *SimCLR/SimCLR-v2* [6, 7], which shows that self-supervised pre-training on ImageNet with a simple contrastive learning framework can generate competitive results on downstream image classification tasks comparing with fully supervised learning. The follow-up work [14] shows that contrastive pre-training can also be applied to supervised settings and further improve the SOTA performance on ImageNet.

Contrastive pre-training is fundamentally a guided clustering process with the objective of learning an embedding space to better separate the samples from different classes, and in turn, the following supervised fine-tuning can be carried out more effectively.

In this paper, we demonstrate that the proposed SCP+SF framework can be effectively applied to medical imaging and boost the performance of the triage screening task. Fig. 1a and 1b (best viewed in color) visualize the sample projections from the proposed contrastive learning module before and after the SCP phase, clearly illustrating the improvement in the separability of the two clusters representing the healthy and at-risk populations. Fig. 1c and 1d are the sample projections from our proposed dual-view model, with the direct supervised learning, and with the proposed SCP+SF training framework, further demonstrating that the SCP+SF results in better clustering quality.

Our experiments show that when trained on our in-house dataset of 134,488 images from 30,487 patients and tested on 2,538 images from 640 patients with biopsy-proven ground truth, our screening models trained with SCP+SF surpass the previously reported SOTA approaches [21, 28, 30] by a large margin, in terms of AUC and specificity/sensitivity.

The main contributions of this paper include: 1) we present a novel framework of SCP+SF, with a carefully designed Siamese contrastive learning module, including details of the network architecture and loss design, and 2) we show that for our mammographic triage screening task, models trained with SCP+SF consistently outperform their directly supervised counterparts and achieve superior performance over previously reported SOTA approaches.

2 Related Work

Contrastive pre-training Most contrastive pre-training works have been conducted within the realm of self-supervised learning on ImageNet data [2, 6, 7, 10, 11, 23, 29, 31], involving different forms of contrastive loss [9]. The recent work of *SimCLR-v2* [7] shows that self-supervised contrastive pre-training can compete with its fully supervised counterpart after fine-tuning on downstream tasks. The work of *SupCon* [14] generalizes the contrastive loss to the supervised setting. In medical imaging, the work of [12] proposes to carry out self-supervised contrastive pre-training at both global and local levels on the Magnetic Resonance Imaging (MRI) dataset before fine-tuning for MRI image segmentation.

Mammographic screening Previous works on deep learning based mammographic screening include two types of triage tasks: 1) identifying the healthy patients to reduce workload [16, 17, 25, 30], and 2) identifying the mammograms with malignant findings [4, 13, 21, 26–28]. We focus on the first task in this paper and treat the BI-RADS 1⁴ category mammograms as healthy/normal.

Most of the above screening methods take the direct supervised learning approach, except for [28], which pre-trained the screening model on a large amount of data with BI-RADS labels before fine-tuning it with biopsy ground truth. However, this pre-training phase is *not* based on the contrastive learning principle, therefore different from our approach.

In particular, the approach in [30] is a classic single-view based method, and the approaches in [21, 28] represent the latest SOTA multi-view mammographic screening methods. All three approaches have been tested on large-scale datasets. Therefore they are selected for comparison with our proposed method in Sec. 5.

3 Method

SCP+SF framework The overall architecture of the SCP+SF framework is illustrated in Fig. 2. The *Siamese contrastive learning module* is designed to carry out the SCP phase, and the resulting Siamese encoders are then transferred to the *single-view learning module* and the *dual-view learning module* to continue the SF phase, respectively, as shown by the magenta arrows in Fig. 2. We further elaborate on both phases in the following subsections.

SCP phase The SCP phase is carried out by the Siamese contrastive learning module, which consists of a Siamese encoding block and a Siamese projection block. In the encoding block, one pair of the input mammographic images are simultaneously fed into the shared-weight encoders. The encoded features are then projected into a lower dimensional space by max-pooling and 1×1 conv operations before flattened into two 1-dimensional vectors. The 1-D vectors are further reduced to 2×1 output vectors through fully connected layers and sigmoid operation, representing the likelihood for each class.

⁴Details regarding the BI-RADS standard can be found in [8].

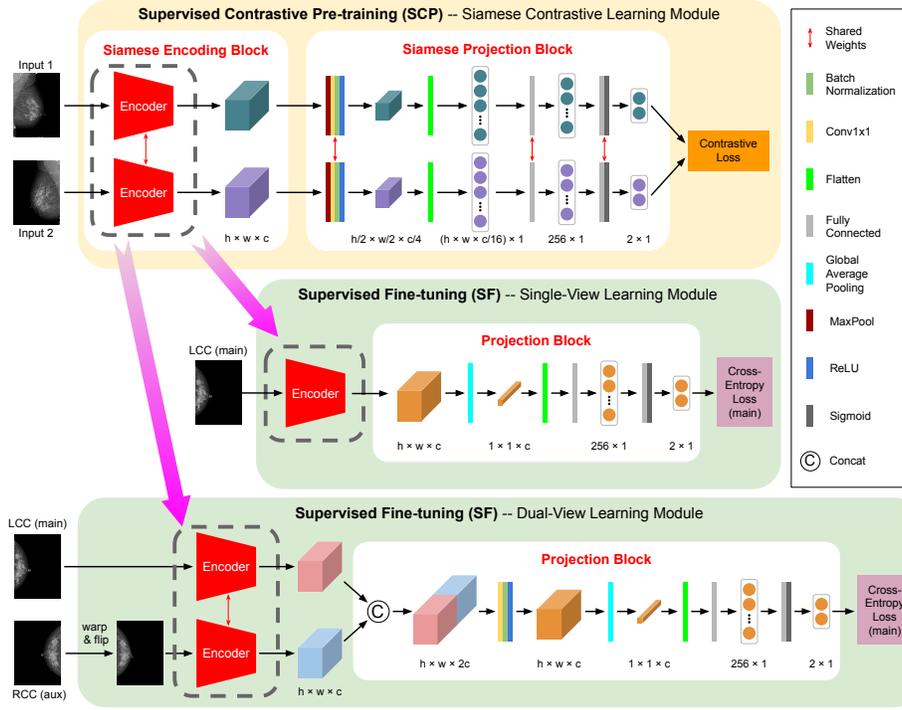


Fig. 2. The architecture of SCP+SF framework, consisting of the Siamese contrastive learning module and SF modules for both the single-view and the dual-view model.

The contrastive loss [9] is designed to draw the samples from the same class closer and separate the samples from different classes farther apart in the projected space. Given a pair of input images (I, I'), we use the regular L2 distance in the loss function and set $margin$ as 1:

$$L(I, I') = \begin{cases} D^2 & \text{if } l_I = l_{I'} \\ \max(0, margin - D)^2 & \text{if } l_I \neq l_{I'}, \end{cases} \quad (1)$$

where

$$D = \|P_{\text{sia}}(E_{\text{sia}}(I)) - P_{\text{sia}}(E_{\text{sia}}(I'))\|_{L_2}, \quad (2)$$

and $E_{\text{sia}}(\cdot)$ and $P_{\text{sia}}(\cdot)$ denote the Siamese encoder block and nonlinear projection block, respectively; l_I and $l_{I'}$ indicate the corresponding BI-RADS labels. The loss for a batch of N image pairs can be simply defined as $\mathcal{L}_{\text{batch}} = \sum_{i=1}^N L(I_i, I'_i)$.

Other types of loss function (e.g., inner product based) as in [6, 14] are also experimented and compared in Sec 5. The SCP phase is completed once the training of this module ends.

SF phase Each mammogram typically includes four views, the left and right craniocaudal (LCC/RCC), and mediolateral-oblique (LMLO/RMLO), and the triage screening model can take one or multiple views as input.

Single-view model The *Single-view learning module* in Fig. 2 illustrates the network architecture of the single-view model. During the SF phase, its encoder block is directly transferred from the Siamese encoding block trained in SCP phase and kept intact, while the projection block after the encoder is fine-tuned based on the regular cross-entropy loss.

Dual-view model In practice, radiologists routinely identify the abnormalities through bilateral analysis of mammography image pairs (i.e., LCC/RCC, or LMLO/RMLO). Therefore we also experimented with the bilateral views as the input for a dual-view model, in addition to the single-view model.

The *Dual-view learning module* in the SF phase comprises a dual-view based input structure, a Siamese encoder, and a projection block, as shown in Fig. 2. Since our screening model output is for each image, we designate one image of the bilateral pair as the main input, and the other image serves as the auxiliary input. For the example shown in Fig. 2, the LCC view is the main input, and the RCC view from the same patient serves as the auxiliary input. The RCC input will first be registered and warped according to the LCC view before being fed into the shared-weight pre-trained encoder, in tandem with the LCC view. The output encoded features are then concatenated before being projected into a lower dimension and further reduced to a 2×1 vector. Similar to the single-view model, the encoder block of the dual-view model is directly transferred from the SCP phase and fixed during the SF phase.

Sample selection strategy During supervised contrastive learning, a batch of images is first randomly selected from the training set, and then the positive and negative pairs are identified according to the sample labels within this selected batch [14]. Limited by the affordable batch size, we experimented with two slightly different sampling strategies. One method is random sampling, where the training batch includes N pairs of positive and negative image pairs directly sampled from *the entire training set* with the corresponding labels. Since our dual-view model takes input from a pair of images from the same patient, we also experimented with a patient-constrained sampling method, where each randomly sampled positive or negative pair must come from *the same patient*.

4 Experiment Design

Our triage screening task aims to identify normal mammograms (BI-RADS 1) with near-perfect accuracy in physical screening scenario. Patients with any suspicious regions in the breasts should not be screened out as normal patients. This task can be further defined as a binary classification problem of BI-RADS 1 (normal/healthy) vs. other BI-RADS categories (abnormal). The majority of screening mammograms belongs to BI-RADS 1, and thus screen them off can assist radiologists to reduce their workload.

Table 1. Number of mammography images in each BI-RADS category.

BI-RADS	Abnormal						<i>Subtotal</i>	Normal	<i>Total</i>
	2	3	0	4	5	6			
Train & Val	35290	14508	1570	3288	732	174	55562	78926	134488
Test	728	226	136	88	6	8	1192	1346	2538

Datasets Our data for training and validation is collected from three collaborative hospitals at distinct geographical locations using Siemens and Giotto equipment in accordance with the ACR standard (American College of Radiology) and dated from 2011 to 2018⁵. This dataset contains both image data and diagnosis reports that are all from screening exams. It includes 30,487 patients, among which 13,931 patients have at least one breast diagnosed as abnormal (other than BI-RADS 1), and 16,556 patients have both breasts diagnosed as normal (BI-RADS 1). Our test set includes 640 patients collected within 31 consecutive days (March 2019) from one of those three hospitals, among which 405 patients have at least one breast diagnosed as abnormal, and 235 patients have both breasts diagnosed as normal. Mammograms in test set come with biopsy proven malignancy results. Table 1 shows the number of images in each BI-RADS category for these two datasets. For abnormal cases, the BI-RADS categories are listed by increasing risk level, where BI-RADS 0 is often regarded as between BI-RADS 3 and 4 by radiologists [3, 22]. In the literature [21, 28, 30], mammographic screening models have been developed and tested only with large-scale private datasets conforming to the realistic patient distribution. Public datasets like DDSM [18] have BI-RADS 1 patients’ data marked as ‘normal’. But they account for less than 30% of the complete dataset, of which the data distribution is not consistent with screening mammography scenarios. Other public mammography datasets either contain no BI-RADS 1 patients’ data or only have diagnostic mammograms (when a screening mammogram does show an abnormality, a diagnostic mammogram may be needed). Since no similar public dataset is available, we follow the convention and perform experiments on our own large-scale datasets.

Implementation details⁶ The dataset is split into the training and validation sets by 8:1 ratio. All input images are resized to 1008×800 and retain the original aspect ratio. The SCP and SF phases share the following training parameter settings. The initial learning rate is 1×10^{-5} with 4 warming-up steps and reduced to 1×10^{-6} after 100 epochs. Adam is used [15], with a weight decay of 5×10^{-4} . Two NVIDIA V100 GPUs (16G memory each) are used, and the batch size for contrastive learning is set to 6 due to the computation limit. The model training normally completes within 300 epochs. Our final model’s runtime is less than 3 seconds on our machine. One limitation is the model does

⁵This retrospective case-control study was approved by the ethics review and institutional review board, which waived the requirement for individual informed consent.

⁶All implementation is with python3.7 and pytorch 1.1.0.

Table 2. Performance comparison of different approaches on the triage screening task. *SCP+SF(R)* refers to SCP with random sample selection and *SCP+SF(P)* indicates SCP with patient-constrained sample selection as in Sec. 3. *SV*, *DV* and *4V* stand for single-view, dual-view and 4-view method, respectively. The last column shows the number of incorrectly screened abnormal images with a breakdown according to their BI-RADS levels (in the order of 2, 3, 0, 4, 5, 6). The 95% confidence intervals (CI) are shown in the square brackets.

Method	AUC	Sensitivity = 20%	
		Specificity	# of <i>incorrectly</i> screened out images out of 1192 abnormal images
[30] SV (2019)	0.8438 [.8418, .8462]	0.9723 [.9687, .9753]	33 (10, 8, 12, 2, 0, 1)
[28] 4V (2019)	0.8702 [.8676, .8734]	0.9773 [.9749, .9797]	27 (9, 8, 9, 1, 0, 0)
[21] 4V (2020)	0.8617 [.8544, .8696]	0.9765 [.9737, .9793]	28 (8, 12, 7, 1, 0, 0)
SV	0.8349 [.8329, .8373]	0.9715 [.9691, .9741]	34 (7, 9, 14, 2, 1, 1)
SV SCP+SF(R)	0.8518 [.8492, .8550]	0.9757 [.9743, .9771]	29 (7, 9, 11, 2, 0, 0)
SV SCP+SF(P)	0.8429 [.8368, .8490]	0.9748 [.9712, .9784]	30 (10, 6, 11, 2, 0, 1)
DV	0.8554 [.8528, .8580]	0.9765 [.9747, .9783]	28 (7, 8, 12, 1, 0, 0)
DV SCP+SF(R)	0.8805 [.8746, .8864]	0.9799 [.9780, .9818]	24 (9, 5, 10, 0, 0, 0)
DV SCP+SF(P)	0.9040 [.9001, .9079]	0.9832 [.9816, .9850]	20 (7, 5, 8, 0, 0, 0)

require GPU to run, and it has not been fully tested on data outside our collaborators. For comparison purposes, we re-implemented methods as in [21, 28, 30] and applied them to our datasets.

Evaluation metric Since the goal is to screen out a portion of normal mammograms with near-perfect accuracy, we set the sensitivity (recall rate of normal images) at 20%, which is commonly used in clinical studies for mammogram triage screening [30], and compare the specificity rate (percentage of correctly classified abnormal images) of different approaches. In addition, AUC is used to compare the overall performance of the classification models.

5 Experimental Results

Table 2 shows the performance comparison of the triage screening task with different approaches, where Yala [30], Wu [28] and McKinney [21] are all previous SOTA models, alongside our proposed single-view and dual-view models with different SCP+SF training strategies. We also scrutinize the number of abnormal images that are incorrectly screened out from the 1,192 abnormal images in the test set, and further break it down according to the BI-RADS level.

Effectiveness of SCP+SF As shown in Table 2, for both single-view and dual-view models, the SCP+SF framework effectively improves the overall performance of the models, including the AUC and the specificity at given sensitivity (20%). In turn, the number of total incorrectly screened abnormal images

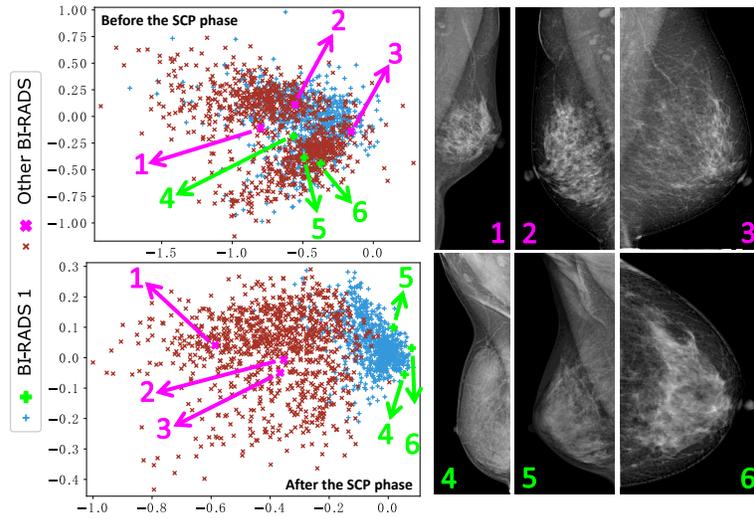


Fig. 3. Visualization of the normal (BI-RADS 1) and abnormal (other BI-RADS) sample projections from the test set data before (top-left) and after (bottom-left) the SCP phase, along with 6 samples and their corresponding images.

is reduced. In addition, for the dual-view model, the SCP+SF framework can completely remove the error made for BI-RADS 4,5,6 images, which is critical in practice since those images often correspond to higher cancer risk. In general, most misclassified abnormal mammograms are due to small scattered benign calcifications, which are occasionally misclassified into BI-RADS 1 category. For the single-view model, the SCP+SF framework can also reduce the error for BI-RADS 5,6 images to near zero. We further confirm from the separate biopsy reports that the incorrectly screened images from the dual-view methods with SCP+SF do not include any malignant findings.

Compared with the previous SOTA approaches, our single-view model with SCP+SF and random sample selection generates the best single-view performance. Our dual-view model trained with SCP+SF and patient-constrained sample selection generates the best result overall, outperforming Wu [28] and McKinney [21], which are both 4-view based models. At higher sensitivity, such as 0.8, our SCP+SF with patient-constrained sample selection method improves the specificity from 0.805 to 0.858 comparing to our vanilla dual view model.

Fig. 3 shows that the normal and abnormal sample projections from the Siamese contrastive learning module on the test data are much better clustered after the SCP phase, further illustrating its efficacy. Three sample images from each class are also given.

Regarding sample selection strategy, the patient-constrained sampling further improves the dual-view model over the random sampling method, while random sampling is slightly better than the patient-constrained sampling for the single-view model, both are consistent with our expectation.

Table 3. Ablation study on different encoders and contrastive loss.

Loss	Inner Product		L2 Loss	
Encoder	ResNet-22	ResNet-34	ResNet-50	ResNet-22
AUC	0.9031	0.9013	0.8993	0.9040
Specificity (Sen.=20%)	98.23%	98.15%	98.15%	98.32%

Ablation study on SCP Table 3 shows the ablation study results. For the backbone, there is no significant difference in terms of the size of ResNet on our task, and ResNet-22 (as in [28]) is selected to serve our encoder. For the loss function, the inner product distance gives a comparable result (slightly worse) as the L2 distance we use in the contrastive loss.

6 Conclusions

We present a novel framework of Supervised Contrastive Pre-training followed by Supervised Fine-tuning (SCP+SF) for mammographic triage screening task. Our experiments with a total of 137,026 images have demonstrated that the SCP+SF framework substantially improved the final model performance, comparing with the direct supervised training. Superior results have also been achieved in comparison with previously reported SOTA approaches. One limitation of this study is that we only applied regular L2 distance for contrastive loss calculation. We plan to experiment with other loss functions, such as triplet loss [5] and magnet loss [24] formulations. We also plan to apply the SCP+SF approach to other medical imaging classification tasks in the future.

References

1. MQSA national statistics, <https://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsa-national-statistics> 1
2. Bachman, P., Hjelm, R.D., Buchwalter, W.: Learning representations by maximizing mutual information across views. In: Advances in Neural Information Processing Systems. vol. 32, pp. 15535–15545. Curran Associates, Inc. (2019) 3
3. Castells, X., Torá-Rocamora, I., Posso, M., Román, M., Vernet-Tomas, M., Rodríguez-Arana, A., Domingo, L., Vidal, C., Baré, M., Ferrer, J., et al.: Risk of breast cancer in women with false-positive results according to mammographic features. *Radiology* **280**(2), 379–386 (2016) 6
4. de Cea, M.V.S., Diedrich, K., Bakalo, R., Ness, L., Richmond, D.: Multi-task learning for detection and classification of cancer in screening mammography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 241–250. Springer (2020) 1, 3
5. Chechik, G., Sharma, V., Shalit, U., Bengio, S.: Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research* **11**(3) (2010) 9
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. vol. 119, pp. 1597–1607. PRML (2020) 2, 3, 4

7. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.E.: Big self-supervised models are strong semi-supervised learners. In: *Advances in Neural Information Processing Systems*. vol. 33, pp. 22243–22255. Curran Associates, Inc. (2020) [2](#), [3](#)
8. D’Orsi, C.: 2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. American College of Radiology (2014) [3](#)
9. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. vol. 2, pp. 1735–1742 (2006) [3](#), [4](#)
10. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 9729–9738 (2020) [2](#), [3](#)
11. Hjelm, R.D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Trischler, A., Bengio, Y.: Learning deep representations by mutual information estimation and maximization. In: *International Conference on Learning Representations* (2019) [2](#), [3](#)
12. Karani, K.C.E.E.N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: *Advances in Neural Information Processing Systems*. vol. 33. Curran Associates, Inc. (2020) [3](#)
13. Khan, H.N., Shahid, A.R., Raza, B., Dar, A.H., Alquhayz, H.: Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. *IEEE Access* **7**, 165724–165733 (2019) [3](#)
14. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. In: *Advances in Neural Information Processing Systems*. vol. 33, pp. 18661–18673. Curran Associates, Inc. (2020) [2](#), [3](#), [4](#), [5](#)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *International Conference on Learning Representations* (2015) [6](#)
16. Kontos, D., Conant, E.F.: Can AI Help Make Screening Mammography “Lean”? *Radiology* **293**(1), 47–48 (2019) [1](#), [3](#)
17. Lång, K., Dustler, M., Dahlblom, V., Åkesson, A., Andersson, I., Zackrisson, S.: Identifying normal mammograms in a large screening population using artificial intelligence. *European Radiology* **31**(3), 1687–1692 (2021) [3](#)
18. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data* **4**, 170177 (2017) [6](#)
19. Lehman, C.D., Arao, R.F., Sprague, B.L., Lee, J.M., Buist, D.S., Kerlikowske, K., Henderson, L.M., Onega, T., Tosteson, A.N., Rauscher, G.H., et al.: National performance benchmarks for modern screening digital mammography: update from the breast cancer surveillance consortium. *Radiology* **283**(1), 49–58 (2017) [1](#)
20. Lehman, C.D.: Artificial Intelligence to Support Independent Assessment of Screening Mammograms—The Time Has Come. *JAMA oncology* **6**(10), 1588–1589 (2020) [1](#)
21. McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., et al.: International evaluation of an ai system for breast cancer screening. *Nature* **577**(7788), 89–94 (2020) [1](#), [2](#), [3](#), [6](#), [7](#), [8](#)
22. Nelson, H.D., O’Meara, E.S., Kerlikowske, K., Balch, S., Miglioretti, D.: Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. *Annals of internal medicine* **164**(4), 226–235 (2016) [6](#)
23. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018) [3](#)

24. Rippel, O., Paluri, M., Dollar, P., Bourdev, L.: Metric learning with adaptive density discrimination. arXiv preprint arXiv:1511.05939 (2015) [9](#)
25. Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Teuwen, J., Broeders, M., Genaro, G., Clauser, P., Helbich, T.H., Chevalier, M., Mertelmeier, T.: Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *European radiology* **29**(9), 4825–4832 (2019) [1](#), [3](#)
26. Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W.: Deep learning to improve breast cancer detection on screening mammography. *Scientific reports* **9**(1), 1–12 (2019) [3](#)
27. Sun, L., Wang, J., Hu, Z., Xu, Y., Cui, Z.: Multi-view convolutional neural networks for mammographic image classification. *IEEE Access* **7**, 126273–126282 (2019) [3](#)
28. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzebski, S., Févry, T., Katsnelson, J., Kim, E., et al.: Deep neural networks improve radiologists’ performance in breast cancer screening. *IEEE transactions on medical imaging* **39**(4), 1184–1194 (2019) [2](#), [3](#), [6](#), [7](#), [8](#), [9](#)
29. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3733–3742 (2018) [2](#), [3](#)
30. Yala, A., Schuster, T., Miles, R., Barzilay, R., Lehman, C.: A deep learning model to triage screening mammograms: a simulation study. *Radiology* **293**(1), 38–46 (2019) [2](#), [3](#), [6](#), [7](#)
31. Zhuang, C., Zhai, A.L., Yamins, D.: Local aggregation for unsupervised learning of visual embeddings. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 6002–6012 (2019) [3](#)