Intrinsic subtype classification of breast lesions on mammograms by contrastive learning

Muramatsu, Chisako, Oiwa, Mikinao, Kawasaki, Tomonori, Fujita, Hiroshi

Chisako Muramatsu, Mikinao Oiwa, Tomonori Kawasaki, Hiroshi Fujita, "Intrinsic subtype classification of breast lesions on mammograms by contrastive learning," Proc. SPIE 12033, Medical Imaging 2022: Computer-Aided Diagnosis, 120331S (4 April 2022); doi: 10.1117/12.2613173

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Intrinsic subtype classification of breast lesions on mammograms by contrastive learning

Chisako Muramatsu[a], Mikinao Oiwa[b], Tomonori Kawasaki[c], Hiroshi Fujita[d]
[a]Faculty of Data Science, Shiga University,
1-1-1 Banba, Hikone, Shiga 522-8522, Japan
[b]Department of Radiology, Nagoya Medical Center,
4-1-1 Sannomaru, Naka-ku, Nagoya, Aichi 460-0001, Japan
[c]Department of Diagnostic Pathology, Saitama Medical University International Medical Center,
1397-1 Yamane, Hidaka, Saitama 350-1298, Japan
[d]Department of Electrical, Electronics and Computer Engineering, Faculty of Engineering,
Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan

ABSTRACT

Periodic breast cancer screening with mammography is considered effective in decreasing breast cancer mortality. Once cancer is found, the best treatment is selected based on the characteristic of cancer. In this study, we investigated a method to classify breast cancer lesions into four molecular subtypes to assist diagnosis and treatment planning. Because of a limited number of samples and imbalanced types, the lesions were classified based on the similarities of samples using a contrastive learning. The convolutional neural network (CNN) was trained by self-supervised method using paired views of the same lesions with contrastive loss. The subtype was determined by k-nearest neighbor classifier using deep features obtained by the trained network. The proposed model was tested using 385 cases by a 4-fold cross validation. The results are compared with CNN models without and with pretraining. The result indicates the potential usefulness of the proposed method. The computerized subtype classification may support a prompt treatment planning and proper patient care.

1. PURPOSE

Mammography is considered an effective screening exam for reducing breast cancer mortality. When a lesion is found, the patient is sent to further diagnostic testing including ultrasonography and MRI. Radiologists make diagnosis by taking all the diagnostic testing results into account and assess if there are no conflicting findings. Although diagnosis of mammograms can be difficult, it may be advantageous to predict possible pathology in advance.

When a cancer is found, the best treatment is selected based on the cancer characteristics, including pathologic types, histological grade, and intrinsic subtypes [1-3]. The use of hormonal therapy and chemotherapy can be determined by the intrinsic subtypes. In this study, we investigated a computerized image analysis method for classification of breast lesions on mammograms into four intrinsic subtypes, including luminal-A type, luminal-B type, HER2 type, and triple-negative/basal-like type. The subtype classification can be useful in assisting radiologists for prompt treatment planning and prediction of prognosis.

There are only a limited number of studies investigating the correlation between subtypes and diagnostic image findings on mammography, ultrasound, and MRI [4-6]. Son et al. [7] proposed a machine learning method to predict molecular subtypes using radiomic features on breast tomosynthesis. Based on the manual contours of lesions, various features including texture features were determined, and classification of luminal, HER2, and triple-negative types was performed using elastic-net.

Although subtype classification on mammograms is extremely difficult, even by expert radiologists, we investigated the possibility of AI-based methods to predict breast lesion subtypes and use of contrastive learning to tackle with extremely unbalanced samples.

Medical Imaging 2022: Computer-Aided Diagnosis, edited by Karen Drukker, Khan M. Iftekharuddin, Hongbing Lu, Maciej A. Mazurowski, Chisako Muramatsu, Ravi K. Samala, Proc. of SPIE
Vol. 12033, 120331S · © 2022 SPIE · 1605-7422 · doi: 10.1117/12.2613173

2. MATERIALS AND METHODS

2.1 Mammography datasets

Digital mammograms used in this study were obtained at Nagoya Medical Center, Japan. All the lesions were diagnosed cancer based on biopsy or surgery. Only one lesion per patient was used in this study. A total of 385 lesions, consisting of 131 luminal-A type, 172 luminal-B type, 47 triple-negative, and 35 HER2 type, were included. Table 1 shows the major mammographic findings of these lesions. Note that many of the lesions have mixture of findings. No findings includes the lesions found by other modalities such as breast tomosynthesis and breast ultrasound. Using rough outlines of lesions provided by a radiologist, a rectangular region of interest was cropped. Each patch was made to a square region by zero padding and then resized to 300 x 300 pixels before input to the networks.

Because of a limited number of cases, another dataset including 162 benign mass lesions and 133 malignant mass lesions obtained at the same institution using various digital mammography units were used for pretraining of the model.

Table 1. Numbers of lesions in intrinsic subtypes and their major mammographic findings

	Luminal-A	Luminal-B	Triple-negative	HER2
Mass	66	80	28	12
Microcalcifications	15	30	4	12
Architectural distortion	15	24	4	1
Focal asymmetric density	20	18	8	8
No findings	15	20	3	2
Total	131	172	47	35

2.2 Proposed model

The number of samples used in this study was limited. In addition, they are highly imbalanced as shown in Table 1. Therefore, the convolutional neural network (CNN) was trained for similarities of lesions by contrastive learning. The contrastive learning was based on SimCLR [8], which is the self-supervised method using augmented samples. In this study, pairs of images obtained from the cranio-caudal (CC) view and mediolateral oblique (MLO) view of the same lesions were treated as positive pairs instead of the augmented images of the same sample. Therefore, it is in a way also a self-supervised training. All the other pairs in a batch were considered as negative pairs as in SimCLR method. This pretraining was performed using 105 pairs (210 images) from the pretraining dataset. Unpaired samples (lesions only clearly depicted in one view) were not used. The contrastive loss in [8] was employed for the network training.

The base network is EfficientNetB0 [9], and a full connection layer with 128 outputs and ReLU activation function is added after the global average pooling layer. The network was trained up to 300 epochs. The framework was Keras and optimizer was Adam with a learning rate of 0.0003.

The lesion subtypes were predicted using k-nearest neighbor (kNN) classifier. The test cases were first input to the trained model, and output 128 deep features were employed for the kNN. The k parameters of 3, 5, and 7 without and with distance weight were tested. The proposed model was evaluated by 4-fold cross validation. The 385 cases were divided into four groups by stratified random sampling.

For comparison, subtype classification using (1) CNN with no pretraining, (2) CNN pretrained with benign and malignant samples, (3) CNN pretrained with imageNet, and (4) CNN pretrained by contrastive learning (fine tuning instead of kNN classifier) were investigated. Figure 1 shows the schematic overview. In pretraining for benign and malignant classification, 268 and 27 samples from the pretraining dataset were used for training and validation, respectively.

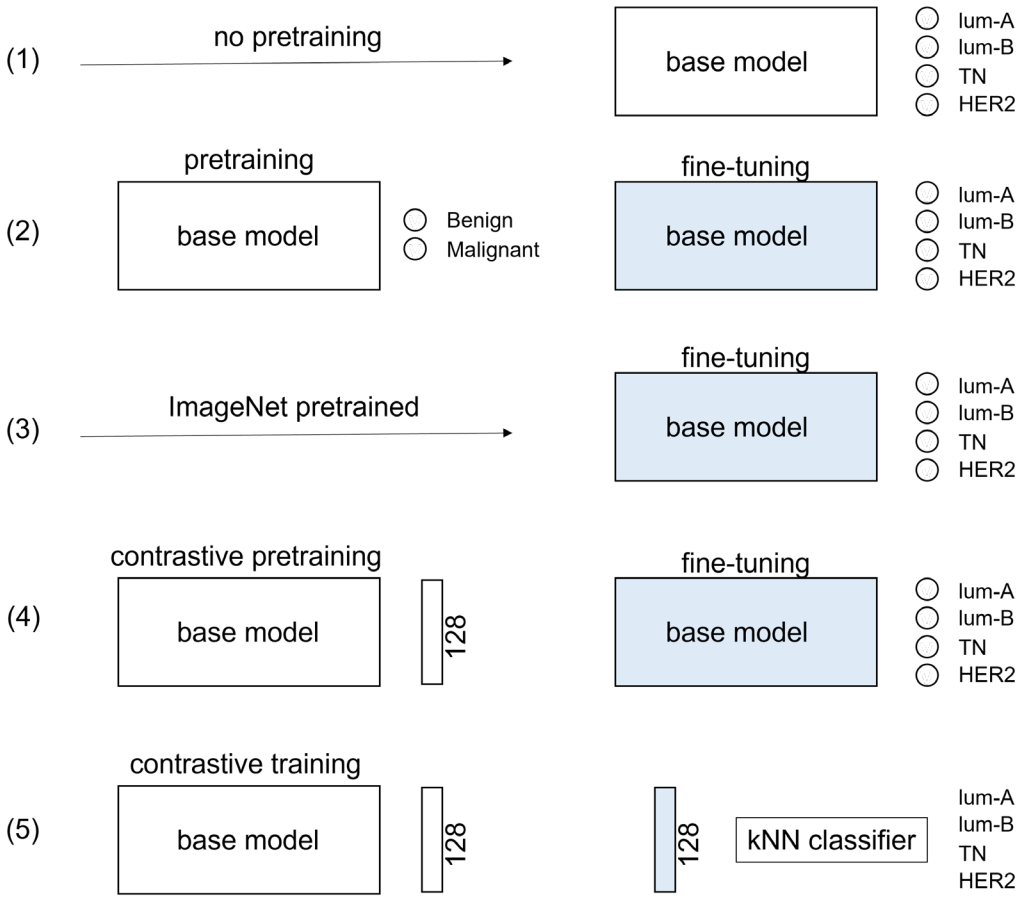

Fig 1. Overview of the proposed and comparison training schemes. (1) CNN model without pretraining, (2) CNN model pretrained for classification of benign and malignant masses, (3) ImageNet pretrained CNN model, (4) CNN model pretrained by contrastive learning, and (5) proposed kNN model using deep features by contrastive learning.

2.3 Evaluation

The test data used in this study is highly unbalanced dataset. Therefore, high accuracy can be obtained by classifying all images to either of two luminal types. We evaluated the proposed model in terms of macro-averaged F1 score as well as accuracy. F1 score was determined for each class and then averaged.

3. RESULTS

The average F1 score was highest by the proposed method. Although accuracy and F1 score are very low because of the difficulty of the task, performance was slightly improved using contrastive learning and kNN compared with the other models. The accuracies and F1 scores for the tested models are listed in Table 2. With the proposed model, k=5 and use of uniform weight in kNN classifier provided the best result, although difference was very small using the distance weight. For the CNN models, data augmentation to balance the number of samples in 4 classes was also examined. However, it degraded the results in methods (1), (2) and (3). Using the model with no pretraining, all cases were classified to either luminal-A or luminal-B types. By the proposed method, some of the HER2 and triple-negative cases were correctly classified. Table 3 shows the confusion matrices for the tested methods.

Table 2. Accuracy and average F1 scores of the tested methods

Method	Accuracy	F1 score
(1) no pretraining	0.40	0.22
(2) pretrained with ben/mal	0.36	0.25
(3) ImageNet-pretrained	0.33	0.27
(4) contrastive pretraining	0.37	0.23
(5) contrastive training + kNN	0.39	0.30

Table 3. Confusion matrices for the methods

(1)	lumA	lumB	TN	HER2
lumA	65	66	0	0
lumB	84	88	0	0
TN	21	26	0	0
HER2	21	14	0	0

(2)	lumA	lumB	TN	HER2
lumA	31	71	28	1
lumB	37	97	36	2
TN	3	35	9	0
HER2	3	22	9	1

(3)	lumA	lumB	TN	HER2
lumA	62	46	20	3
lumB	98	51	18	5
TN	24	10	11	2
HER2	19	10	3	3

(4)	lumA	lumB	TN	HER2
lumA	55	66	10	0
lumB	75	86	11	0
TN	20	24	3	0
HER2	18	17	0	0

(5)	lumA	lumB	TN	HER2
lumA	55	62	2	12
lumB	74	85	1	12
TN	17	23	6	1
HER2	13	18	0	4

4. DISCUSSION

In this study, we investigated image analysis methods to classify breast cancers on mammograms into four intrinsic subtypes. The classification of intrinsic subtypes on mammography is a very difficult task. In addition, our dataset includes cases with various findings, such as mass, microcalcifications, architectural distortion, and those originally with no specific findings on mammograms which lesions were found by other imaging modalities. Figure 2 shows examples of tomo-found lesions. Moreover, the test dataset is extremely imbalanced.

Son et al. [7] proposed a computerized method for classification of three molecular subtypes, namely, luminal A and B, HER2, and triple-negative, on breast tomosynthesis. Using elastic-net and hand-crafted features, they achieved relatively high classification performance. However, the number of tested cases was very small (50 luminal, 12 triple-negative, and 9 HER2 types), and the method requires the manual contours of the lesions. They used equal numbers of cases in training the model while remaining the unbalanced data for testing, which could be effective in achieving good performance.

Although classification accuracy was very low, our result indicates the potential usefulness of contrastive learning combined with the kNN classifier to provide improved average F1 score. The self-supervised contrastive learning method proposed by Chen et al. [8] achieved high classification performance using small dataset that is equivalent to supervised

training with a large dataset. We also tested the method proposed in [8] which is self-supervised learning with data augmentation and fine-tuning with classification head. The accuracy was higher with 0.47; however, all cases were classified into luminal A or B types (F1 score of 0.25). In this study, we used CC-MLO view pairs of the same lesions as positive samples in contrastive learning, since they are natural similar pairs. However, the number of paired samples were small, and they include benign pairs while test cases were all cancer cases. In the self-supervised contrastive learning, lesions with the same subtypes could be trained as negative (dissimilar) pairs. Supervised contrastive learning with cancer cases may improve the pretraining and classification result.

We investigated the use of contrastive learning to possibly solve the imbalanced sample problem. Although F1 score was slightly improved, the result was insufficient. Oversampling with data augmentation was not effective in this study. The use of kNN classifier was somewhat effective in predicting minority types. Other methods for imbalanced data problem must be examined and may be combined with the proposed method.

In this study, mammographic regions of interest were used for subtype prediction. Use of other modality images such as ultrasound images together with mammograms may improve classification performance. Multi-task training including pathologic types, histological grades, and invasiveness may be helpful in predicting intrinsic subtypes.

Classification of breast lesions into four intrinsic subtypes was investigated for possible assistance in treatment planning. Classification performance must be greatly improved to support radiologists in prompt diagnosis and precision medicine.

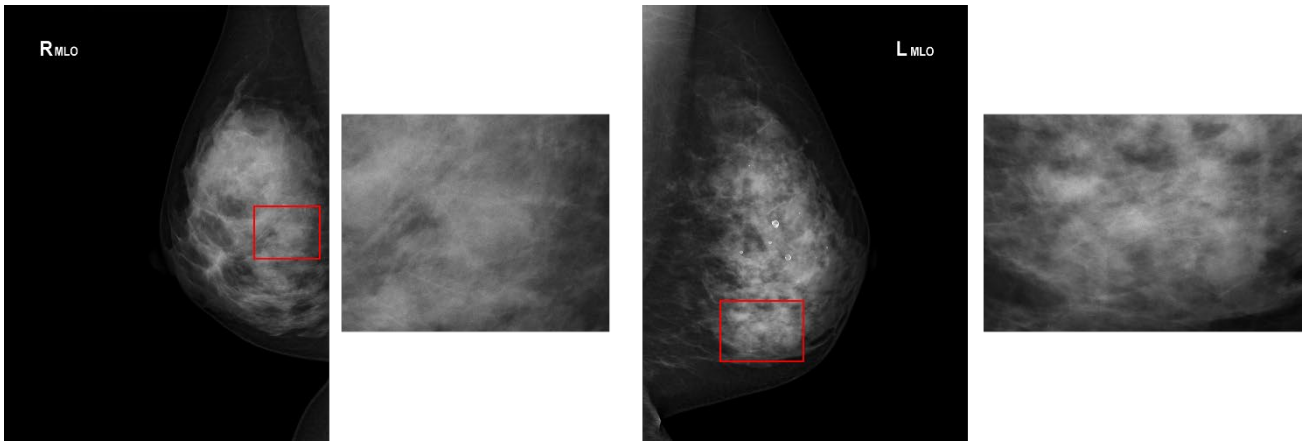

Fig 2. Lesions originally with no finding on mammography and found by breast tomosynthesis. Left: luminal-B type and Right: triple-negative type.

ACKNOWLEDGMENT

This study was partly supported by a Grant-in-Aid for Scientific Research (C) (No. 20K08131) by Japan Society for the Promotion of Science.

REFERENCES

[1] Perou, C. M., Sorlie, T., Elise, M. B., Rijin van de, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, Hilde., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borrensen-Dale, A. L., Bwown, P. O., Botstein, D., "Molecular portraints of human breast tumours," Nature 406, 747–752 (2000).

[2] Goldhirsch, A., Wood W. C., Coates, A. S., Gelber, R. D., Thurlimann, B., Senn, H. J., "Strategies for subtypes - deeling with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011," Annals Oncol 22, 1736-1747 (2011).

[3] Wolff, A. C, McShane, L. M., Hammond E. H., Allison, K. H., Fitzgibbons, P., Press, M. F., Harvey, B. E., Mangu, P. B., Bartlett, J. M. S., Hanna, W., Bilous, M., Ellis, I. O., Dowsett, M., Jenkins, R. B., Spears, P. A., Vance, G. H., Viale, G., "Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical

Oncology/College of American Pathologists Clinical Practice Guideline Focused Update," Arch Pathol Lab Med 142, 1364-1382 (2018).
[4] Uematsu, T., Kasami, M., Yuen, S., "Triple-negative breast cancer: correlation between MR imaging and pathologic findings," Radiology 250, 638-647 (2009).
[5] Celebi, F., Pilance, K. N., Ordu, C., Agacayak, F., Alco, G., Ilgun, S., Sarsenov, D., Erdogan, Z., Ozmen, V., "The role of ultrasonographic findings to predict molecular subtype, histologic grade, and hormone receptor status of breast cancer," Diagn Interv Radiol 21, 448-453 (2015).
[6] Wu, M., Ma, J., "Association between imaging characteristics and different molecular subtypes of breast cancer," Acad Radiol 24, 426-434 (2017).
[7] Son, J., Lee, S. E., Kim, E. K., Kim, S., "Prediction of breast cancer molecular subtypes using radiomics signatures of synthetic mammography from digital breast tomosynthesis," Sci Rep 10, 21566 (2020).
[8] Chen, T., Kornblith, S., Norouzi, M., Hinton, G., "A simple framework for contrastive learning of visual representations," Proc ICML 119 (2020).
[9] Tan, M., Le, Q. V., "EfficientNet: Rethinking model scaling for convolutional neural networks," Proc ICML 97 (2019).