Sensors and Materials, Vol. 36, No. 5 (2024) 1951–1957
MYU Tokyo

S & M 3646

Mammographic Breast Composition Classification Using Swin Transformer Network

Kuen-Jang Tsai,[1,2] Wei-Cheng Yeh,[3,4] Cheng-Yi Kao,[3] Ming-Wei Lin,[5,6] Chao-Ming Hung,[2] Hung-Ying Chi,[7] Cheng-Yu Yeh,[7*] and Shaw-Hwa Hwang[8]

[1]Department of Chemical Engineering, and Institute of Biotechnology, I-Shou University, Kaohsiung 82455, Taiwan
[2]Department of General Surgery, E-Da Cancer Hospital, Kaohsiung 82445, Taiwan
[3]Department of Radiology, E-Da Cancer Hospital/I-Shou University, Kaohsiung 82445, Taiwan
[4]Department of Information Engineering, I-Shou University, Kaohsiung 82445, Taiwan
[5]Department of Medical Research, E-Da Hospital/E-Da Cancer Hospital, I-Shou University, Kaohsiung 82445, Taiwan
[6]Department of Nursing, College of Medicine, I-Shou University, Kaohsiung 82445, Taiwan
[7]Department of Electrical Engineering, National Chin-Yi University of Technology, 57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan
[8]Department of Electronics and Electrical Engineering, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan

(Received December 15, 2023; accepted May 10, 2024)

Keywords: screening mammography, breast imaging reporting and data system (BI-RADS), breast composition, image classification, Swin Transformer, deep learning

Breast cancer is a prevalent global health concern and the most commonly diagnosed cancer in women. Mammography, a well-established and widely used screening tool, has greatly contributed to early breast cancer detection. However, understanding mammographic breast composition is also crucial for refining the risk assessment of breast cancer beyond identifying lesions. In contrast to previous studies, we adopt an exploratory approach by using the Swin Transformer, a foundation model for image classification, to classify the four-category breast density. Leveraging this foundation, we fine-tune the model with a small set of mammograms for the purpose of making advancements in breast density classification. This study is experimentally validated to achieve an overall accuracy of 74.96% in the four-category breast density classification, which is a comparable performance to recent counterparts.

1. Introduction

Breast cancer is a pervasive global health issue and the most commonly diagnosed cancer among women.(1) Early detection remains pivotal in improving outcomes and reducing mortality rates associated with this disease. Mammography, a well-established and widely utilized screening tool, has significantly contributed to the early detection of breast cancer.(2) However, the interpretation of mammographic images extends beyond the mere identification of lesions. Understanding the intricacies of breast tissue composition, known as mammographic breast

*Corresponding author: e-mail: cy.yeh@ncut.edu.tw
https://doi.org/10.18494/SAM4826

ISSN 0914-4935

composition, has emerged as a critical aspect in refining breast cancer risk assessment and screening strategies.

Mammographic breast composition refers to the proportion and distribution of various tissues within the breast as visualized through mammography. Central to its clinical relevance is the recognition that breast density, a key component of breast composition, serves as a crucial indicator for breast cancer risk. Studies have consistently demonstrated an increased risk associated with higher breast density,[3,4] making it imperative to comprehensively evaluate breast composition for effective risk stratification.

The American College of Radiology's breast imaging reporting and data system (BI-RADS)[5] provides a standardized framework for characterizing breast composition. The four BI-RADS categories (A to D) offer a systematic approach to communicating the density and composition of breast tissues. These categories, ranging from predominantly fatty to extremely dense, guide radiologists in reporting findings consistently and enable healthcare providers to make informed decisions regarding further assessments.

Traditionally, the analysis of breast composition heavily relies on the subjective interpretation of radiologists, demanding significant time and effort for accurate assessment. However, with the advent of artificial intelligence (AI), algorithms based on deep learning have demonstrated exceptional capabilities in handling extensive medical imaging data, extracting intricate patterns, and providing valuable insights. These capabilities hold the potential to enhance the accuracy of breast composition analysis,[6–8] enabling radiologists to make interpretations more rapidly and improving overall diagnostic efficiency.

In recent times, numerous groups have proposed solutions for breast density classification based on deep learning, yielding promising results.[9–13] Diverging from previous studies, we adopt an exploratory approach using the Swin Transformer,[14,15] a foundation model for image classification, to classify the four-category breast density. The Swin Transformer has also demonstrated exceptional performance in the field of image classification. The advantage of employing the Swin Transformer lies in its pretraining on an extensive dataset of images, allowing it to extract various features from this vast image repository. Leveraging this foundation, we fine-tune the model with a small set of mammograms, aiming to make advancements in breast density classification.

2. Materials

Table 1 summarizes the description of breast composition, that is, breast density, in the BI-RADS mammography lexicon.[5] There are four categories in breast composition. Category A corresponds to breasts with low density, where the majority of tissue is composed of fat. Category B represents breasts with some fibroglandular tissue dispersed among fatty tissue. Category C indicates breasts with increased density owing to a higher proportion of fibroglandular tissue, which potentially makes detection more challenging. Category D comprises breasts with the highest density, where fibroglandular tissue predominates, increasing the difficulty of mammographic detection and breast cancer screening.

Table 1
Description of breast composition in BI-RADS mammography lexicon.

Category	Description
A	Entirely fatty
B	Scattered areas of fibroglandular density
C	Heterogeneously dense, which may obscure masses
D	Extremely dense, which lowers sensitivity

In this work, the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM)[16] is used to train the presented model for breast composition classification. CBIS-DDSM is a public dataset designed for medical imaging research, specifically focused on breast imaging. It can be used for the performance evaluation in computer-aided diagnosis and detection systems research in mammography. The dataset is composed of 3101 mammograms of 1566 participants, including 425, 1207, 952, and 517 images for categories A to D in breast composition, respectively. Table 2 gives the numbers of training and test data used in this study.

3. Methodology and Model

We aim to classify breast composition into four categories using the Swin Transformer model,[14,15] which can assist radiologists in making faster interpretations and improves the overall diagnostic efficiency. The Swin Transformer, proposed by Microsoft Research Asia, is a foundation model based on the transformer architecture. It has demonstrated outstanding performance in image classification and capably serves as a general-purpose backbone for computer vision.

The advantage of using the Swin Transformer lies in its pretraining on an extensive dataset of images, enabling it to extract various features from this vast image repository. On the basis of this concept, we fine-tune the model with a small set of mammograms, aiming to make advancements in breast density classification.

In this paper, we adopt the base model of the Swin Transformer, called Swin-B, with the architecture depicted in Fig. 1. First of all, the Swin-B model employed in this study was pretrained on the ImageNet-22K dataset.[17] Next, the input image size is specified as 224 × 224, that is $H = W = 224$ in Fig. 1. Subsequently, in Fig. 1, $C = 128$, and the model comprises a total of four stages. The Swin Transformer Block in each stage is successively concatenated with a different number, i.e., 2, 2, 18, and 2, respectively. The number of connections is indicated below in each Swin Transformer Block.

Finally, a categorical cross-entropy loss function was used to train the model with a batch size of 64 and an epoch of 1000, and an AdamW optimizer[18] with a learning rate of 0.00001 was also used to improve the training performance. Table 3 lists the development environment of this work. As a consequence, the presented model occupies 331 MB of memory.

Table 2
Numbers of training and test data in each composition category.

Breast composition category	Number of training data	Number of test data
A	358	67
B	945	262
C	751	201
D	404	113
Sum	2458	643

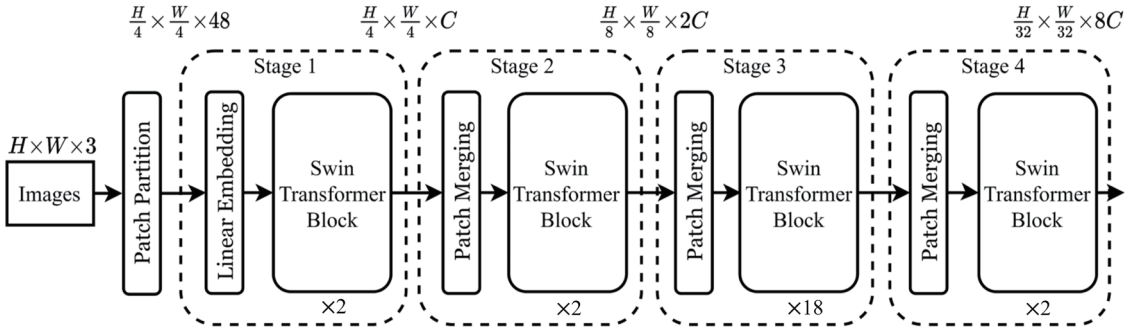

Fig. 1. Architecture of Swin-B model.

Table 3
Development environment.

Programing language	Python
Library	Pytorch 1.13.1, CUDA 11.7, numpy, opencv-python, matplotlib, etc.
Hardware	PC (Windows 10 64-bit, AMD Ryzen Thredripper PRO 3955WX 16-Cores 3.9 GHz CPU, 256 GB RAM), graphics card (NVIDIA RTX A6000)

4. Experimental Results

A confusion matrix for four-category classification and four performance metrics for each class, including the sensitivity, specificity, precision, and *F*1-*score*, were conducted to evaluate the model performance. Then, the overall accuracy was obtained. The four performance metrics—sensitivity (*Sen*), specificity (*Spe*), precision (*Pre*), and *F*1-*score*—are respectively given by

$$Sen_i = TP_i / (TP_i + FN_i), \tag{1}$$

$$Spe_i = TN_i / (TN_i + FP_i), \tag{2}$$

$$Pre_i = TP_i / (TP_i + FP_i), \tag{3}$$

$$F1\text{-}score_i = 2 \times (Pre_i \times Sen_i) / (Pre_i + Sen_i), \tag{4}$$

where $1 \leq i \leq Cnum = 4$ and is used to represent that a predictive result is classified as the ith element of the breast composition category: {A, B, C, D}. True positive (*TP*) and false positive (*FP*) are used to represent whether a predictive result is accurately classified or misclassified in the ith case, respectively. Likewise, true negative (*TN*) and false negative (*FN*) represent whether a predictive result is accurately classified or misidentified in a case other than the ith case, respectively. Finally, the overall accuracy is given by

$$Accuracy = \frac{\sum_{i=1}^{Cnum} TP_i}{Tnum}, \tag{5}$$

where *Tnum* is the number of test data.

Performance testing was conducted using the 643 test data values tabulated in Table 2, the confusion matrix in Fig. 2, and the performance metrics in Table 4. The model has an overall accuracy of 74.96% for four-level breast density classification. The *F*1-*score* had its highest (79.35%) and lowest (53.45%) values in categories B and A, respectively. A further examination of Fig. 2 reveals that the breast density was misclassified as one level higher or lower in most misclassified cases.

Finally, Table 5 presents the overall accuracies in this study and three recently published counterparts. Despite differences in the test datasets, the accuracy of this study, at 74.96%, is notably higher than that in Ref. 10 and is comparable to those of the methods proposed in Refs. 11 and 12. This provides evidence that implementing the foundation model with the Swin Transformer for breast composition classification is indeed successful and feasible.

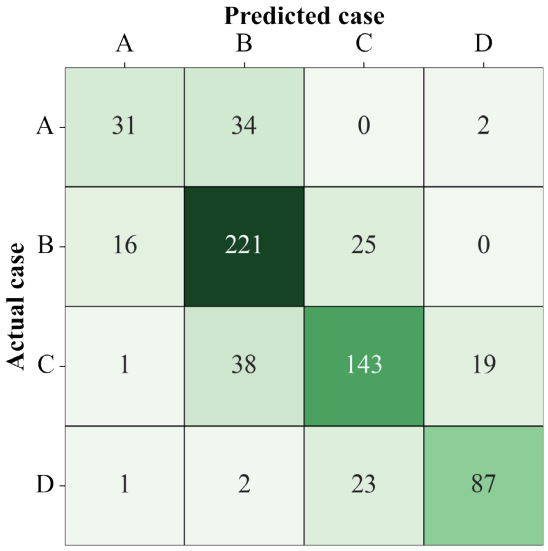

Fig. 2. (Color online) Confusion matrix for performance analysis.

Table 4
Performance metrics in the proposed approach.

Breast density	*Sen* (%)	*Spe* (%)	*Pre* (%)	*F*1-*score* (%)
A	46.27	96.88	63.27	53.45
B	84.35	80.58	74.92	79.35
C	71.14	89.14	74.87	72.96
D	76.99	96.04	80.56	78.73
Accuracy (%)		74.96		

Table 5
Overall accuracies in this study and counterparts.

Reference	Dataset	Accuracy (%)
This study	CBIS-DDSM	74.96
Ref. 10 (2019)	DDSM	68
Ref. 11 (2022)	Private	75
Ref. 12 (2022)	DDSM	74.94 (single view)

5. Conclusions

Unlike previous literature, in this paper, we present the use of the Swin Transformer foundation model to perform the task of four-category breast density classification. Given its pretraining on an extensive dataset of images, the Swin Transformer has the advantage of extracting various features from this vast image repository. Consequently, remarkable performance is demonstrated in the field of image classification. Additionally, experimental results confirm that in this study, we achieved an overall accuracy of 74.96% in four-category breast density classification. In the future, a dual view of mammographic breast images will be introduced to build deep learning models with the aim of achieving higher performance than that demonstrated in this study.

References

1 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray: CA-A Cancer J. Clinicians **71** (2021) 209.
2 C. Coleman: Semin. Oncol. Nurs. **33** (2017) 141.
3 M. Román, J. Louro, M. Posso, R. Alcántara, L. Peñalva, M. Sala, J. del Riego, M. Prieto, C. Vidal, M. Sánchez, X. Bargalló, I. Tusquets, and X. Castells: Eur. Radiol. **31** (2021) 4839.
4 F. T. H. Bodewes, A. A. van Asselt, M. D. Dorrius, M. J. W. Greuter, and G. H. de Bock: Breast **66** (2022) 62.
5 A. A. Rao, J. Feneis, C. Lalonde, and H. Ojeda-Fournier: Radiographics **36** (2016) 623.
6 L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb: Comput. Biol. Med. **131** (2021), 104248.
7 S. P. Lester, A. S. Kaur, and S. Vegunta: Oncologist **27** (2022) 548.
8 C. R. Taylor, N. Monga, C. Johnson, J. R. Hawley, and M. Patel: Diagnostics **13** (2023) 2041.
9 A. A. Mohamed, W. A. Berg, H. Peng, R. C. Jankowitz, and S. Wu: Med. Phys. **45** (2018) 314.
10 P. H. Yi, A. Lin, J. Wei, A. C. Yu, H. I. Sair, F. K. Hui, G. D. Hager, and S. C. Harvey: J. Digital Imaging **32** (2019) 565.
11 V. Magni, M. Interlenghi, A. Cozzi, M. Alì, C. Salvatore, A. A. Azzena, D. Capra, S. Carriero, G. Della Pepa, D. Fazzini, G. Granata, C. B. Monti, G. Muscogiuri, G. Pellegrino, S. Schiaffino, I. Castiglioni, S. Papa, and F. Sardanelli: Radiol.: Artif. Intell. **4** (2022) e210199.

12 M. Busaleh, M. Hussain, H. A. Aboalsamh, Fazal-e-Amin, and S. A. Al Sultan: Mathematics **10** (2022) 4610.
13 L. T. Duong, C. Q. Chu, P. T. Nguyen, S. T. Nguyen, and B. Q. Tran: Appl. Soft Comput. **134** (2023) 109974.
14 Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo: Proc. IEEE/CVF Int. Conf. Computer Vision (IEEE, 2021) 9992. https://doi.org/10.1109/ICCV48922.2021.00986
15 Microsoft Swin Transformer Code: https://github.com/microsoft/Swin-Transformer (accessed September 2023).
16 R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin: Sci. Data **4** (2017) 170177.
17 J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and Li Fei-Fei: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2009) 248. https://doi.org/10.1109/CVPR.2009.5206848
18 I. Loshchilov and F. Hutter: arXiv (2019). https://doi.org/10.48550/arXiv.1711.05101