



Impact of Downsampling Size and Interpretation Methods on Diagnostic Accuracy in Deep Learning Model for Breast Cancer Using Digital Breast Tomosynthesis Images

Ryusei Inamori,¹ Tomofumi Kaneno,¹ Ken Oba,² Eichi Takaya,^{1,3}
Daisuke Hirahara,⁴ Tomoya Kobayashi,^{1,3} Kurara Kawaguchi,¹ Maki Adachi,¹
Daiki Shimokawa,¹ Kengo Takahashi,¹ Hiroko Tsunoda² and Takuya Ueda¹

¹Tohoku University Graduate School of Medicine, Department of Clinical Imaging, Aoba-ku, Sendai, Miyagi, Japan

²St. Luke's International Hospital, Department of Radiology, Chuo-ku, Tokyo, Japan

³Tohoku University Hospital, AI Lab, Aoba-ku, Sendai, Miyagi, Japan

⁴Department of AI Research Lab, Harada Academy, Kagoshima, Kagoshima, Japan

While deep learning (DL) models have shown promise in breast cancer diagnosis using digital breast tomosynthesis (DBT) images, the impact of varying matrix sizes and image interpolation methods on diagnostic accuracy remains unclear. Understanding these effects is essential to optimize preprocessing steps for DL models, which can lead to more efficient training processes, improved diagnostic accuracy, and better utilization of computational resources. Our institutional review board approved this retrospective study and waived the requirement for informed consent from the patients. In this study, 499 patients (29-90 years old, mean age 50.5 years) who underwent breast tomosynthesis were included. We performed downsampling to 256×256 , 128×128 , 64×64 , and 32×32 using five image interpolation methods: Nearest (NN), Bilinear (BL), Bicubic (BC), Hamming (HM), and Lanczos (LC). The diagnostic accuracy of the DL model was assessed by mean AUC with its 95% confidence interval (CI). DL models with downsampled images to 256×256 pixels using the LC interpolation method showed a significantly lower AUC than the original 512×512 pixels model. This decrease was also observed with the 128×128 pixels DL models using HM and LC methods. All interpolation methods showed a significant decrease in AUC for the 64×64 and 32×32 pixels DL models. Our results highlight the significant impact of downsampling size and interpolation methods on the diagnostic performance of DL models. Understanding these effects is essential for optimizing preprocessing steps, which can enhance the accuracy and reliability of breast cancer diagnosis using DBT images.

Keywords: breast cancer diagnosis; deep learning; digital breast tomosynthesis; downsampling; preprocessing
Tohoku J. Exp. Med., 2025 January, 265 (1), 1-6.
doi: 10.1620/tjem.2024.J071

Introduction

Breast cancer is the most common cancer affecting women worldwide, and its incidence and mortality rates are expected to increase (Anastasiadi et al. 2017; Harbeck et al. 2017). Mammography has proven to be an effective screening tool for early detection of breast cancer (Harbeck et al. 2017). The sensitivity and specificity for diagnosing breast cancer using mammography by physicians are reported to

be 86.9% and 88.9%, respectively (Lehman et al. 2017). Hamashima et al. (2015) reported that a meta-analysis of five randomized controlled trials showed a 25% reduction in mortality with mammography alone.

Recently, digital breast tomosynthesis (DBT), a new and advanced breast imaging technique, has been applied in clinical practice. DBT allows for volumetric reconstruction of the entire breast from several two-dimensional projections obtained at different X-ray tube angles (Nelson et al.

Received January 24, 2024; revised and accepted July 11, 2024; J-STAGE Advance online publication July 25, 2024

Correspondence: Takuya Ueda, M.D., Ph.D., Tohoku University Graduate School of Medicine, Department of Diagnostic Radiology, 2-1 Seiryomachi, Aoba-ku, Sendai, Miyagi 980-8575, Japan.
e-mail: takuya.ueda.d3@tohoku.ac.jp

©2025 Tohoku University Medical Press. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC-BY-NC-ND 4.0). Anyone may download, reuse, copy, reprint, or distribute the article without modifications or adaptations for non-profit purposes if they cite the original authors and source properly.
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

2009). Thinly sliced images acquired by DBT improve sensitivity and specificity compared to mammographic images by reducing the overlap between breast tissue and lesions, especially in dense breast tissue (Anastasiadi et al. 2017). Previous research has shown that DL models based on FFDM and DBT images reduce the workload of physicians by 29.7% without compromising the quality of results (Mendel et al. 2019; Raya-Povedano et al. 2021).

Deep learning (DL) has been applied in various medical fields, such as speech recognition, visual object recognition, and object detection (LeCun et al. 2015). Although the application of DL has been boosted by the development of GPUs, large datasets, and advanced algorithms (Rawat and Wang 2017), the limited size of GPU memory creates a tradeoff between batch size and image matrix size used to train DL models (Sabottke and Spieler 2020). Batch size is the number of images used for each parameter update in stochastic gradient descent (Kandel and Castelli 2020). Kandel and Castelli (2020) reported that batch size affects the accuracy of DL models and the time taken until convergence. Other research reported that extremely small batch sizes slow convergence during training and degrade performance during inference (Yan et al. 2020; Lin 2022). Optimizing batch size is essential for achieving high-performing DL models. However, batch size needs to be reduced due to the limited size of GPU memory when images with large matrix sizes are used to train DL models (Sabottke and Spieler 2020).

Downsampling is commonly applied as a preprocessing step for deep learning models that work with image datasets with large matrix sizes. The application of downsampling sufficiently increases the batch size during DL training, so appropriate downsampled images improve the performance of DL models. There are several kinds of interpolation methods to reduce the matrix size of images, such as Nearest (Lehmann et al. 1999), Bilinear (Lehmann et al. 1999), and Bicubic (Keys 1981). Since the matrix size of medical images is commonly much larger than that of images in other fields, such as natural images (Willeminck et al. 2020), the batch size for medical images is more severely limited by GPU memory compared to natural images. Sabottke and Spieler (2020) found that reducing the matrix size through downsampling did not significantly affect the performance of DL models for diagnosing chest radiographs. DL models trained on downsampled chest radiographic images performed comparably to those trained on larger images (Sabottke and Spieler 2020). Hirahara et al. (2021) reported that the type of interpolation method for downsampling chest radiographs affects the performance of DL models. As the image matrix size of FFDM and DBT images in breast cancer imaging is much larger than other medical images, including chest radiography, the benefit of downsampling may be even greater for DL models for breast cancer imaging than for chest radiography (Lehman et al. 2017). The purpose of our study is to investigate the impact of different matrix sizes and various image interpo-

lation methods on the diagnostic performance of DL models for breast cancer classification using DBT images. Our study provides insights into optimal preprocessing techniques for DL models, ensuring that diagnostic accuracy is maintained while maximizing computational efficiency. This knowledge may contribute to the development of more robust and efficient DL models and facilitate their practical application in medical image processing. By identifying the most effective downsampling sizes and interpolation methods, we aim to enhance the overall performance and reliability of DL models.

Materials and Methods

Patient enrollment

The institutional review board approved this retrospective study and waived the requisite to obtain the informed consent from the patients. Fig. 1 shows flowchart of patient enrollment for this study. A total of 499 patients (mean age of 50.5 years, ranging from 29 to 90 years) who admitted and underwent DBT between March 1, 2019 and August 31, 2019, were enrolled in this study. This study used bilateral mediolateral oblique (MLO) views of DBT imaging from 978 breasts of 499 patients. Out of the 978 breasts, we excluded 331 breasts on the basis of lack of bilateral imaging (20 breasts), gynecomastia (10 breasts), post-operation (36 breasts), with metal clips placed after biopsy (5 breasts), and cases with inaccurate annotation due to inaccurate localization by DBT (260 breasts). As a result, DBT image dataset of 647 breasts were enrolled in this study including 170 breasts with pathologically confirmed breast cancers and 477 breasts with benign lesions (198 breasts) or normal breast tissue (279 breasts).

The determination of normal or benign lesions were those confirmed by histopathology or those for which ultrasound and magnetic resonance imaging.

The DBT dataset was randomly split into two datasets with 80% and 20% for training dataset and test dataset, respectively. To ensure that images from the same patient do not overlap between the training and test datasets, different slice images from the same patient were allocated to either the training or test dataset. Finally, 170 breasts with breast cancer and 477 breasts without breast cancer (198 with benign lesions and 279 normal breasts) were analyzed in this study. Out of the 170 breasts with pathologically confirmed breast cancer, 103 lesions were depicted as a mass, and 71 were identified as calcifications. Some breasts had multiple lesions, and each of these was counted independently. Furthermore, 14 lesions exhibited characteristics of both mass and calcification and were counted in both categories.

Clinical interpretation of tomosynthesis images

All DBT images were acquired on the 3D Mammography System (Hologic, Inc., Bedford, MA). The scanning parameters for the DBT images were as follows: kilovoltage peak ranged from 26 to 45 kV; current from 140

to 200 mA; exposure time from 154 to 489 ms; compression force from 26.7 to 191.4 N; breast thickness from 13 to 108 mm; and absorbed dose from 0.0092 to 0.0688 Gy. The total tomographic angle range was 15° , spanning from -7.5° to 7.5° , consisting of 15 projection views taken at 1° increments. The interslice interval was 1 mm, and the resolution was $70 \mu\text{m} \times 70 \mu\text{m}$ per pixel. All breast lesions were diagnosed and annotated by radiologists with over five years of experience in breast cancer imaging. To study the effect of downsampling on mass visualization, the diameter of the breast cancer masses was investigated. The sizes of 100 breast cancer masses were measured.

Image preprocessing

The images were converted into 16-bit PNG format and cropped to 512×512 pixels, centered at the coordinates of the region annotated by the radiologists. If a cropped image at 512×512 pixels overlapped with the DBT image boundary, it was automatically shifted in parallel to ensure all pixels fit within the image range. For images of normal breast tissue, square areas of 512×512 pixels were randomly selected and cropped, avoiding any breast cancer lesions (Li et al. 2020; Yu et al. 2020). All the original cropped images of 512×512 pixels were then downsampled to 256×256 , 128×128 , 64×64 , and 32×32 pixels using each of the five interpolation methods available in the Python Pillow image processing library. These methods include Nearest (NN) (Lehmann et al. 1999), Bilinear (BL) (Lehmann et al. 1999), Bicubic (BC) (Keys 1981), Hamming (HM) (Harris 1978),

and Lanczos (LC) (Duchon 1979). NN is an algorithm that calculates the pixel value by taking the nearest neighbor point among the four adjacent points (Lehmann et al. 1999). BL is a linear interpolation method that computes the average pixel value based on the distances to the four surrounding points (Lehmann et al. 1999). BC is a cubic interpolation method that interpolates pixels according to the adjacent 16 pixels, fitting the profile of circular diffraction gratings (Keys 1981). HM uses a hamming window for interpolation, where the end values of the window are zero, thus avoiding signal reflection in the spectrum (Harris 1978). LC is characterized by discontinuities at the interval's ends and approximates the sinc filter, with each interpolated value being the weighted sum of two consecutive input samples (Duchon 1979). Fig. 2 illustrates examples of four grades of downsampled images using these five interpolation methods on an original 512×512 pixel image.

Implementation environment

The networks were implemented on a machine equipped with an Intel Core i7-7800X CPU, featuring 6 cores, and an NVIDIA QUADRO RTX 8000 Graphics Processing Unit with 48 GB of memory. The operating system used was Ubuntu 18.04.5 LTS (Long Term Support), Xenial Xerus. All analyses were conducted using Python, version 3.8.2 (Python Software Foundation at <http://www.python.org>). The deep learning framework employed was PyTorch, version 1.5.1.

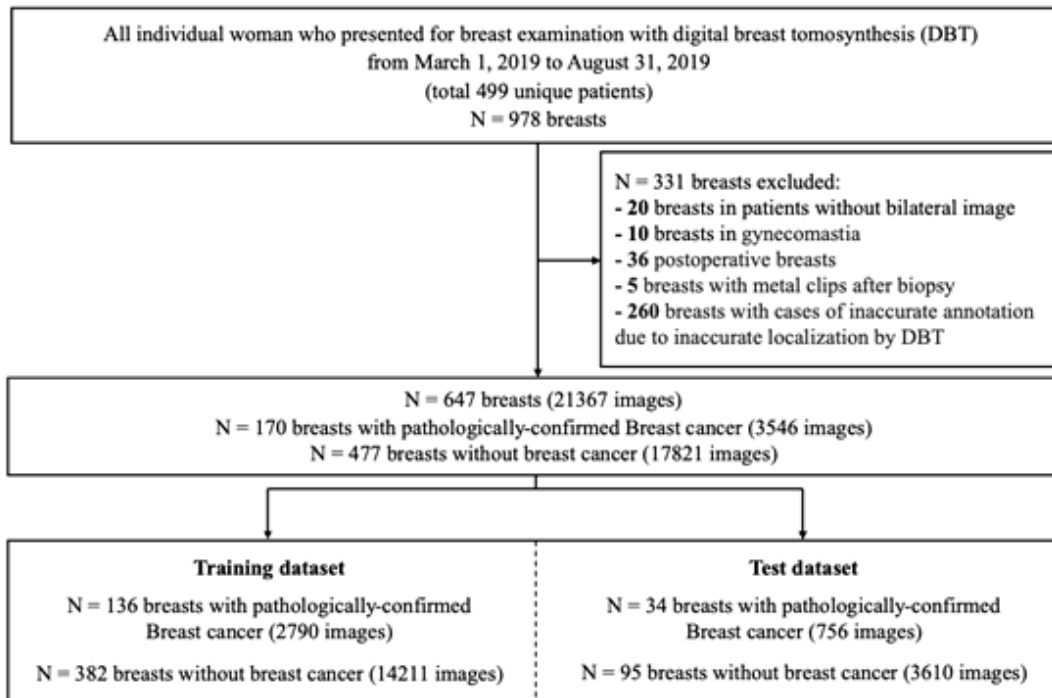


Fig. 1 Flowchart of inclusion and exclusion in the present study. In total, 647 breasts were analyzed in this study.

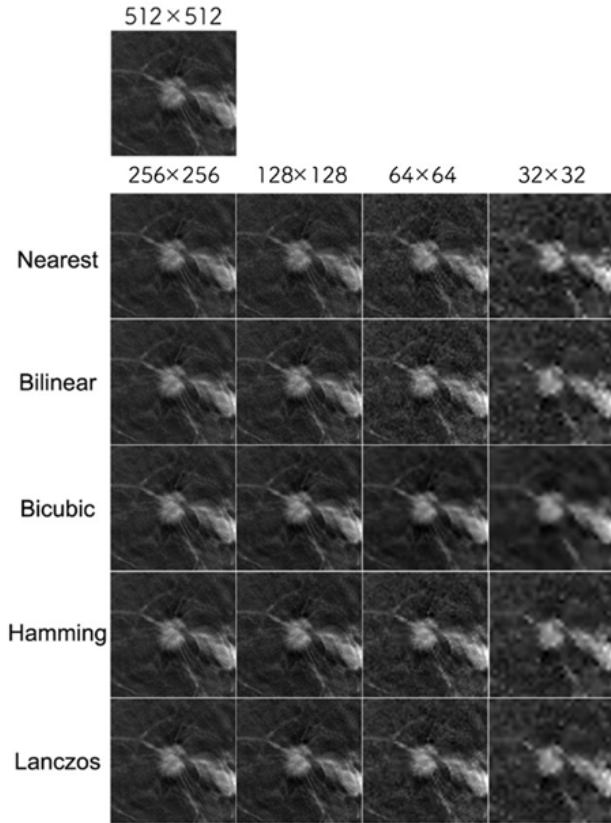


Fig. 2. Interpolation methods for an original 512×512 pixel image. The appearance of the downsampled images varies depending on the type of image interpolation method used.

DL model

A residual neural network (ResNet50), which has been widely applied in various DL models for breast cancer imaging (Yala et al. 2019; Shen et al. 2019), was utilized for the convolutional neural networks (CNN) for binary classification (He et al. 2016). The network weights were initialized based on a model pre-trained on ImageNet, a process known as transfer learning. Adam was selected as the optimizer, and categorical cross-entropy was employed as the loss function (learning rate = 0.01, weight decay = 0.001). The batch size was set at 64, and the number of epochs was set at 100. The DL model underwent training and testing 10 times in one set, saving the probability value of breast cancer for each image in the test set after each run. The model was reinitialized for each iteration of the process.

Assessment of the DL model

The diagnostic accuracy of the DL model was assessed using the mean Area Under the Curve (AUC) along with its 95% confidence interval (CI). To compare the diagnostic accuracy across different matrix sizes, the mean AUC value of the 512×512 dataset was compared with those of other matrix sizes. In this study, the mean AUC values for each

interpolation method at each matrix size were evaluated against the 512×512 matrix dataset. A p-value of less than 0.05 was considered statistically significant. The Student's t-test was employed for statistical significance testing.

Results

The sizes of the 103 breast cancer masses ranged from 5 mm to 87 mm, with a mean of 20.9 mm and a standard deviation of 11.2 mm.

Table 1 and Fig. 3 present the average AUC and the 95% confidence interval (CI) of the AUC for the original DL model with a 512×512 matrix dataset, alongside 20 DL models applying five different image interpolation methods to four datasets with varying matrix sizes. The AUC for the original 512×512 matrix dataset DL model was 0.727, with a 95% CI of 0.712 to 0.742. For the DL model with the downsampled 256×256 matrix dataset, the Lanczos interpolation method showed a significantly lower AUC compared to the original 512×512 matrix dataset DL model, as depicted in Fig. 3. Similarly, for the DL model with the downsampled 128×128 matrix dataset, the Hamming and Lanczos interpolation methods exhibited significantly lower AUC values than the original 512×512 matrix dataset DL model. For the DL models with the downsampled 64×64 and 32×32 matrix datasets, all five interpolation methods demonstrated significantly lower AUC values compared to the original 512×512 matrix dataset DL model.

Discussion

Our results indicate that the diagnostic performance of the DL model varies depending on the degree of downsampling and the choice of interpolation methods. This finding suggests that careful consideration of the appropriate degree of downsampling and the most suitable interpolation methods is crucial when preprocessing images for deep learning models. In determining the appropriate degree of downsampling, the spatial resolution must be considered relative to the size of the target lesion. In our dataset, the mean diameter of breast cancer masses was 21.0 mm, equivalent to approximately 300 pixels in our tomosynthesis images. The size of microcalcifications in breast cancer, ranging from 0.1 to 1 mm (Henrot et al. 2014), corresponds to 1 to 14 pixels in these images. Reducing a 512×512 matrix size image to 64×64 through downsampling reduces the mean diameter of tumors and microcalcifications to approximately 35 and fewer than 4 pixels, respectively. As a result, characteristic morphologies of tumor-forming breast cancer, such as marginal irregularity, and of malignant calcifications may be lost at this level of downsampling, directly impacting the diagnostic performance of the DL model.

The relationship between batch size and image data size is complementary; reducing the matrix size through downsampling allows for an increase in batch size. A larger batch size offers several benefits, including more efficient

Table 1. The average AUC of our DL models in 5 matrix sizes and 5 interpolation methods.

ResNet50 (95%CI) p-value	256 × 256	128 × 128	64 × 64	32 × 32
Nearest	0.728 (0.712-0.737) P = 0.947	0.727 (0.716-0.745) P = 0.972	0.671 (0.658-0.695) P < 0.05	0.629 (0.602-0.647) P < 0.05
Bilinear	0.723 (0.701-0.746) P = 0.830	0.701 (0.685-0.717) P = 0.056	0.690 (0.677-0.704) P < 0.05	0.646 (0.613-0.680) P < 0.05
Bicubic	0.747 (0.730-0.763) P = 0.138	0.703 (0.681-0.724) P = 0.078	0.662 (0.637-0.687) P < 0.05	0.616 (0.579-0.653) P < 0.05
Hamming	0.723 (0.703-0.743) P = 0.786	0.688 (0.673-0.703) P < 0.05	0.676 (0.655-0.696) P < 0.05	0.655 (0.636-0.673) P < 0.05
Lanczos	0.706 (0.693-0.719) P < 0.05	0.688 (0.671-0.705) P < 0.05	0.655 (0.647-0.663) P < 0.05	0.641 (0.609-0.674) P < 0.05

P-value results for the average AUC and 95% confidence interval of AUC, 512×512 obtained from 10 iterations of training and testing in ResNet50 for 4 matrix sizes and 5 interpolation methods. The p-value indicates statistical significance against AUC for the 512×512 dataset. The average AUC with 95% CI in original matrix size (512×512 pixels) was 0.727 (0.712 - 0.742).

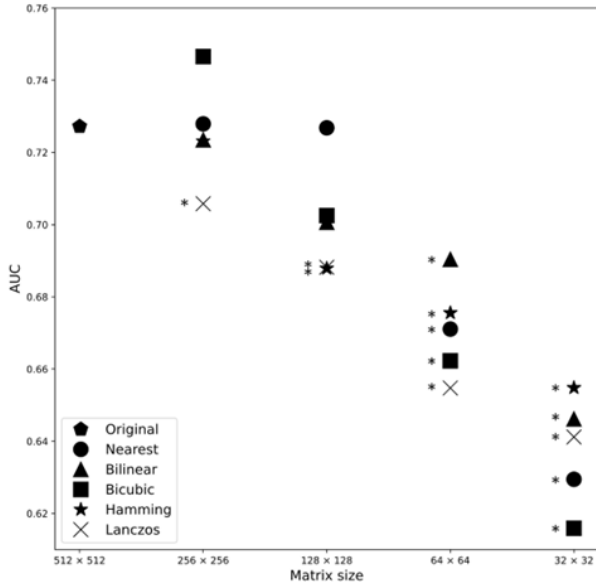


Fig. 3. The average AUC of our DL models in 5 matrix sizes and 5 interpolation methods.

The p-value indicates statistical significance against AUC for the 512×512 dataset. *p < 0.05

hardware resource utilization and the capacity to process more data simultaneously, potentially accelerating training and making the entire process more efficient (Lin 2022). Larger batch sizes provide a more accurate estimate of the gradient, and some research suggests that models trained with larger batch sizes may better generalize from training data to unseen data (Kandel and Castelli 2020). Different

image interpolation methods can also have varying impacts on diagnostic performance.

Our study has several limitations. First, we investigated the diagnostic performance of deep learning models with a fixed batch size. As the difference in batch size affects learning rate and size optimization influences learning rate optimization (Bjorck et al. 2018; Lin 2022), the relationship between appropriate batch size and learning rate in DL warrants further investigation. Second, some benign lesions in our study were diagnosed based on image analysis without follow-up, posing a potential risk of misdiagnosis. Third, our dataset was derived from a single institution's breast tomosynthesis images. To generalize our findings, validation of the DL model with other external, independent datasets is necessary. Fourth, while our study focused on DBT images of breast cancers, the appropriate downsampling size and interpolation method need to be independently evaluated when targeting different types of imaging modalities and diseases.

Conclusion

Our results suggest that careful consideration of the appropriate degree of downsampling and the choice of the most suitable interpolation methods is essential when pre-processing images for deep learning models. Both factors significantly affect the diagnostic performance of the DL model.

Author Contributions

Conceptualization: Ryusei Inamori, Tomofumi Kaneno, Daisuke Hirahara, Takuya Ueda; Methodology: Ryusei Inamori, Eichi Takaya; Investigation: Ken Oba,

Hiroko Tsunoda; Formal analysis: Ryusei Inamori; Validation: Ryusei Inamori, Tomofumi Kaneno; Data Curation: Daiki Shimokawa, Kengo Takahashi, Ken Oba, Kurara Kawaguchi, Maki Adachi, Tomofumi Kaneno, Hiroko Tsunoda; Software: Ryusei Inamori, Eichi Takaya; Resources: Ken Oba, Hiroko Tsunoda; Writing - Original Draft: Ryusei Inamori, Tomofumi Kaneno; Writing - Review & Editing: Ken Oba, Hiroko Tsunoda, Tomoya Kobayashi, Takuya Ueda; Supervision: Hiroko Tsunoda, Takuya Ueda; Funding acquisition: Takuya Ueda

Funding

This work was supported by JST (CREST Grant No. JPMJCR15D1).

Conflict of Interest

The authors declare no conflict of interest.

References

- Anastasiadi, Z., Lianos, G.D., Ignatiadou, E., Harissis, H.V. & Mitsis, M. (2017) Breast cancer in young women: an overview. *Updates Surg.*, **69**, 313-317.
- Bjorck, N., Gomes, C.P., Selman, B., & Weinberger, K.Q. (2018) Understanding Batch Normalization. arXiv:1806.02375.
- Duchon, C.E. (1979) Lanczos Filtering in One and Two Dimensions. *J. Appl. Meteorol. Climatol.*, **18**, 1016-1022.
- Hamashima, C., Ohta, K., Kasahara, Y., Katayama, T., Nakayama, T., Honjo, S. & Ohnuki, K. (2015) A meta-analysis of mammographic screening with and without clinical breast examination. *Cancer Sci.*, **106**, 812-818.
- Harbeck, N. & Gnant, M. (2017) Breast cancer. *The Lancet*, **389**, 1134-1150.
- Harris, F.J. (1978) On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE*, **66**, 51-83.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778.
- Henrot, P., Leroux, A., Barlier, C. & Genin, P. (2014) Breast microcalcifications: the lesions in anatomical pathology. *Diagn. Interv. Imaging*, **95**, 141-152.
- Hirahara, D., Takaya, E., Kadowaki, M., Kobayashi, Y. & Ueda, T. (2021) Effect of the Pixel Interpolation Method for Downsampling Medical Images on Deep Learning Accuracy. *J. Comput. Commun.*, **9**, 150-156.
- Kandel, I. & Castelli, M. (2020) The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, **6**, 312-315.
- Keys, R. (1981) Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust., Speech, Signal Process.*, **29**, 1153-1160.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, **521**, 436-444.
- 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. & Miglioretti, D.L. (2017) National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*, **283**, 49-58.
- Lehmann, T.M., Gonner, C. & Spitzer, K. (1999) Survey: interpolation methods in medical image processing. *IEEE Trans. Med. Imaging*, **18**, 1049-1075.
- Li, X., Qin, G., He, Q., Sun, L., Zeng, H., He, Z., Chen, W., Zhen, X. & Zhou, L. (2020) Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. *Eur. Radiol.*, **30**, 778-788.
- Lin, R. (2022) Analysis on the Selection of the Appropriate Batch Size in CNN Neural Network. In *Proceedings of the 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*, pp. 106-109.
- Mendel, K., Li, H., Sheth, D. & Giger, M. (2019) Transfer Learning From Convolutional Neural Networks for Computer-Aided Diagnosis: A Comparison of Digital Breast Tomosynthesis and Full-Field Digital Mammography. *Acad. Radiol.*, **26**, 735-743.
- Nelson, H.D., Tyne, K., Naik, A., Bougatsos, C., Chan, B.K., Humphrey, L. & Force, U.S.P.S.T. (2009) Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann. Intern. Med.*, **151**, 727-737, W237-742.
- Rawat, W. & Wang, Z. (2017) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.*, **29**, 2352-2449.
- Raya-Povedano, J.L., Romero-Martin, S., Elias-Cabot, E., Gubern-Merida, A., Rodriguez-Ruiz, A. & Alvarez-Benito, M. (2021) AI-based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation. *Radiology*, **300**, 57-65.
- Sabotke, C.F. & Spieler, B.M. (2020) The Effect of Image Resolution on Deep Learning in Radiography. *Radiol. Artif. Intell.*, **2**, e190015.
- Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R. & Sieh, W. (2019) Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci. Rep.*, **9**, 12495.
- Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L. & Lungren, M.P. (2020) Preparing Medical Imaging Data for Machine Learning. *Radiology*, **295**, 4-15.
- Yan, J. & Zhang, X. (2020) Towards Stabilizing Batch Statistics in Backward Propagation of Batch Normalization. arXiv:2001.06838.
- Yala, A., Lehman, C., Schuster, T., Portnoi, T. & Barzilay, R. (2019) A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. *Radiology*, **292**, 60-66.
- Yu, X., Pang, W., Xu, Q. & Liang, M. (2020) Mammographic image classification with deep fusion learning. *Sci. Rep.*, **10**, 14361.