Artificial Intelligence Research and Development
A. Cortés et al. (Eds.)

doi:10.3233/FAIA220348

Breast Tumor Classification in Digital Tomosynthesis Based on Deep Learning Radiomics

Loay HASSAN [a,1], Mohamed ABDEL-NASSER [a,b], Adel SALEH [c] and Domenec PUIG [a]

[a] *Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili,43007 Tarragona, Spain*
[b] *Department of Electrical Engineering, Aswan University, Egypt*
[c] *Gaist Solutions Ltd., Skipton BD23 2TZ, UK*

Abstract. Breast cancer is the most frequently diagnosed cancer in women globally. Early and accurate detection and classification of breast tumors are critical in improving treatment strategies and increasing the patient survival rate. Digital breast tomosynthesis (DBT) is an advanced form of mammography that aids better in the early detection and diagnosis of breast disease. This paper proposes a breast tumor classification method based on analyzing and evaluating the performance of various of the most innovative deep learning classification models in cooperation with a support vector machine (SVM) classifier for a DBT dataset. Specifically, we study the ability to use transfer learning from non-medical images to classify tumors in unseen DBT medical images. In addition, we utilize the fine-tuning technique to improve classification accuracy.

Keywords. Breast Cancer Classification, Digital breast tomosynthesis, Computer vision, Deep learning, Support Vector Machine

1. Introduction

Breast cancer is one of the most deadly deceases affecting females worldwide. The clinical studies have shown that early identification and classification of breast tumors has considerably improved the patient's treatments [1]. Generally, current common modalities for breast cancer screenings include mammography (X-ray images of the breast), breast ultrasound (BUS), thermograms, magnetic resonance imaging (MRI), and digital breast tomosynthesis (DBT) [2,3,4,5,6]. Digital breast tomosynthesis (DBT) is a promising new imaging modality for breast cancer screening that has the potential to overcome the limitations of traditional mammography. Instead of the projectional 2-dimensional images like in mammography, DBT delivers depth information through a practically 3-dimensional structural image of the breast volume (cross-sectional slices) and offers better performance [7]. DBT involves passing the X-ray tube in an arc over a fixed com-

[1]Corresponding Author: Loay Hassan, Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili,43007 Tarragona, Spain; E-mail: loay.abdelrahimosmanhassan@urv.cat.

pressed breast at numerous angles to obtain a series of mammographic images [8]. Computer software subsequently reconstructs these individual mammographic images into a sequence of 3-dimensional and high-resolution slices. These 3-dimensional images notably decrease the influence of dense tissues, which can obscure or make it even harder to identify breast tumors [9].

While DBT can alleviate the limitation of tissue superimposition in mammography by providing superior tissue visibility, misdiagnosis is common, particularly in the early stages of breast cancer. So, a professional pathologist's experience and knowledge are required for a reliable breast cancer diagnosis. Regrettably, examining a more significant number of slices per breast volume raises clinical workflow issues accompanied by the fact that as the number of slices to examine grows, experts' monitoring of findings increases. Therefore, several computer-aided diagnostic (CAD) systems help clinicians identify breast tumors on various breast image modalities. In particular, with DBT images, the performance of the CAD systems will likely be better.

Recently, with the noteworthy development in the deep learning frameworks in many medical field studies [10], researchers have been inspired to investigate the application of deep learning to study the use of deep learning in developing efficient automated CAD systems for the breast cancer classification task [11]. For instance, Chougrad et al. [12] explored the performance of the three recent and powerful state-of-the-art CNN models to predict the correct diagnosis for various mammography breast cancer datasets. The obtained results demonstrate that the explored models are performant and can predict if the tumors are benign or malignant with accuracy ranging from 95 to 98% with different datasets. Adeyinka et al. [13] presented a discriminative fine-tuning deep learning-based method for breast cancer classification on mammograms. Their work aims to perform fine-tuning training using five popular pre-trained CNN models. Unlike traditional fine-tuning, which involves training the entire network again using a specified dataset, discriminative fine-tuning is introduced where they assign different learning rates and momentum to each layer of the network during the training process. The performance of their method is evaluated using the INBreast dataset [14] achieving the highest accuracy of 99% by the DensNet model.

For DBT images, Bevilacqua et al. [15] proposed a supervised deep learning-based normal/abnormal lesions classification method for breast tomosynthesis images. Practically, they compared the performance of two different classification approaches. The first approach is to utilize a shallow artificial neural network (ANN) classifier that takes morphological and textural hand-crafted features like the Grey Level Co-occurrence Matrix (GLCM) as input. The second approach is based on automatically computed features and extracting several sets of features using deeper CNN from DBT images. The final results showed that the second classification approach performs better. With a private DBT dataset, they obtained an accuracy of 92% with the VGG network. Samala et al. [16] developed a multi-stage deep learning framework for classifying malignant and benign tumors in DBT images. The main idea of that work is to study the effectiveness of the transfer learning approach when a pre-trained CNN model on non-medical images is first fine-tuned to a related task in the medical imaging domain before being fine-tuned to the target task in an attempt to overcome the lack of large training data. In practice, the proposed framework consists of two stages. In the first stage, they utilized the AlexNet model trained by more than a million non-medical images from the ImageNet dataset to be fine-tuned with less than 3000 patch images extracted from mammograms. Then, they used the fine-tuned model from the first stage to further fine-tuned with less than 1500 patches from DBT images in the second stage. The experimental results evaluated on a private DBT dataset showed that the accuracy of multi-stage transfer learning is improved by 6% over single-stage transfer learning.

Zhang et al. [17] presented breast normal/abnormal lesions classification method in DBT images where a traditional 2-D deep CNN model is operated on the whole volume of 3-D DBT images, regardless of the number of slices. The fundamental idea behind their work is that instead of only using small tumor patches, they focus on full-image classification. Specifically, for z-slices of the DBT image, every three consecutive slices are stacked as a three-channel image input to the feature extractor network. Then, they generated a feature map by pooling the features extracted for the binary classification. The experimental results evaluated using a private clinical DBT dataset showed that classification performed best using the AlexNet model as feature extractor with MaxPooling for feature fusion.

Although various breast tumor classification methods in DBT are proposed, most of them are limited to classifying normal and abnormal lesions. Also, most of these proposed studies are evaluated using a private DBT dataset. Therefore, automated benign/malignant tumor classification of the breast in DBT still faces several challenges due to the lack of public available DBT datasets. In this paper, based entirely on the only publicly available DBT dataset, we present benign/malignant tumor classification using a support vector machine (SVM) classifier and deep learning classifier for DBT images based on radiomics extracted from well-known CNN classification models. We explore the ability to use transfer learning from non-medical images to classify tumors in DBT images. Alongside, the fine-tuning approach is applied to improve the classifier's efficiency.

The remainder of this paper is designed as follows. Section 2 describes the proposed classification method. The experimental results and discussion are presented in section 3. Section 4 concludes the paper.

2. Methodology

Figure 1 shows the proposed breast tumor classification method. The key elements of the proposed method are data preparation, deep learning-based feature (Radiomics) extraction, and classification heads. Below, we describe the proposed method in detail.

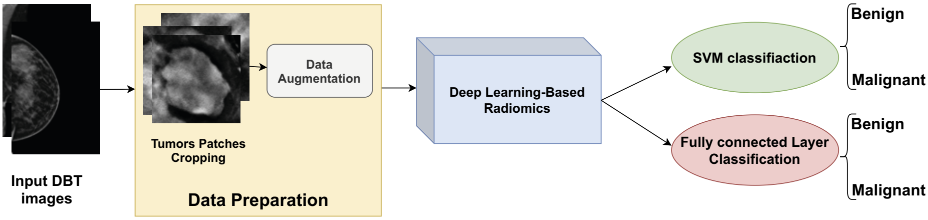

Figure 1. Breast tumor classification in DBT images based on deep learning-based radiomics.

2.1. Data Preparation

In this study, the data preparation step consists of two stages: the tumor patch cropping stage and the data augmentation stage. It should be noted that we used the only publicly available DBT images dataset named the DBTex challenge dataset [18] which contains 1000 breast tomosynthesis scans from 985 patients. Disastrously, not all images are fully annotated, as out of the 101 patients, only 224 DBT images have annotations leaving us with few usable DBT images.

For the tumor patch cropping stage, in terms of ROIs selection, tumors are center cropped and extracted from the annotated DBT image. The selected ROIs are then all resized into identical dimensions of 224×224, to fit the input of deep learning networks in the feature extraction step. Then, we divide the DBT images (patient-wise) and the corresponding patches into training and testing sets as shown in Table 1.

Regarding data balance, we seek to achieve that the training and the test set involve the equivalent amount of two classes of tumors. Therefore, in the data augmentation stage, the number of training data tumor patches increases (see Table 1). In particular, to balance the benign and malignant number of patches in the training set, we doubled the number by jointly flipping all malignant patches in the training set horizontally and vertically. This eventually results in 120 benign tumor patches and 126 malignant tumor patches, with total patches of 246 in the training set, in addition to 23 benign tumor patches and 23 malignant tumor patch, with a total of 46 in the test set, which achieves a ratio of approximately 80% of the training images and 20% of the test images of the whole set of images.

Table 1. Overview of the DBT dataset.

	No. of patients		**No. of tumor patches**	
	Train	**Test**	**Train**	**Test**
Benign	50	12	120	23
Malignant	27	12	63	23
Total	77	24	183	46
Augmented Data	77	24	**246**	46

2.2. Deep learning-based radiomics

In the breast tumor classification task, each tumor input patch supposes to be classified as benign or malignant. However, in the case of automated machine-based classifiers, this task is much more difficult due to the need to characterize those input images into discriminative features. In contrast to the traditional methods based on hand-crafted feature extraction, deep learning models enable robust and automated feature extraction.

As shown in Figure 2, the radiomics extraction process has been carried out in two phases. In the first phase, a transfer learning approach, classification CNN models pre-trained on non-medical images are used to directly extract features from the medical DBT images to train one of the classifiers in the classification heads. Since these models have been trained on a dataset with many images, for example, trained on the ImageNet dataset, which contains 1.2 million non-medical images for a 1000 class image classification problem, they can extract meaningful features that can be used for direct classifi-

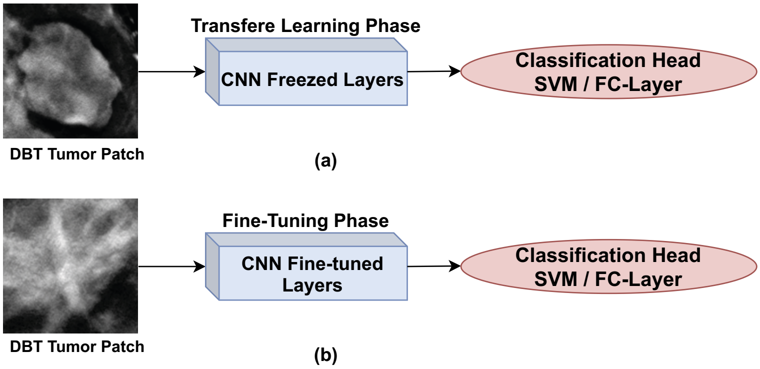

Figure 2. Feature extraction approaches: (a) transfere learning, and (b) fine-tuning.

cation.
In the second phase, a fine-tuning approach, classification CNN models pre-trained on non-medical images are fine-tuned on the training set of 246 DBT images. Applying fine-tuning allows us to utilize the pre-trained networks to the related breast tumor classification task in the DBT medical imaging domain, leading to higher accuracy. For both feature extraction phases, we investigate the performance of various of the most innovative deep learning models inspired by the well-known CNN backbone architectures such as AlexNet [19], VGG [20], ResNet [21], WideResNet [22], SqueezeNet [23] and EfficientNet [24] to extract discriminative features from DBT images.

2.3. Classification heads

Lastly, the generated tumor radiomics from the feature extraction stage can be classified by various machine learning classification algorithms. In this paper, we employ the classification process through two scenarios: end-to-end deep learning classification by a fully connected (FC) layer and an SVM classifier.
In the case of the end-to-end deep learning classification, we modify the classification part of the last FC-layer to the set of classes of our task, i.e., two classes of benign and malignant. Therefore, for the transfer learning approach, we do not need to re-train the entire model, just the final classification part, which will be trained from scratch on top of the pre-trained model. Differently, the fine-tuning approach jointly trains both the newly-modified classifier layers and the whole layers of the base model.
In this study, we also employ the SVM algorithm, a fast and dependable classification algorithm that performs very well with a limited amount of data to analyze, making it suitable for the DBT breast tumor classification task due to the lack of data. Accordingly, whether the radiomics are extracted using transfer learning or fine-tuning approaches, these deep features are used to train the SVM classifier.

2.4. Implementation

To train the end-to-end deep learning classifier for both transfer learning and fine-tuning approaches, we use Adam to optimize the evaluated CNN model with a learning rate of $1e-4$. It should be noted that all models were trained for 30 epochs. On the other hand, to train the SVM classifier, radiomics are extracted first using one of the approaches mentioned above (i.e., transfer learning or fine-tuning), then extracted features are used as input for training and evaluating the SVM classifier. For fair comparisons, the proposed evaluated classifiers were trained using the same 246 DBT images and tested with the same 46 DBT images. All the experiments were performed using the Pytorch framework using a 64-bit Ubuntu operating system with 3.6 GHz intel core i7 with 32GB of RAM and Nvidia RTX3080 with 10GB of video RAM.

3. Experimental Results and Analysis

3.1. Performance evaluation of transfer learning approach

Table 2 presents a quantitative comparison between the SVM classifier and the Deep learning classifier trained with radiomics extracted from the evaluated models (AlexNet, VGG, ResNet, wideResNet, SqueezeNet, and EfficientNet) for the transfer learning approach in terms of percentage accuracy.

As one can see, the SVM classifier trained with radiomics extracted from the pre-trained

Table 2. The performance of the SVM classifier and deep learning classifier with transfer learning approach.

Model	SVM classifier	Deep learning classifier
AlexNet	63.04%	56.52%
VGG19	**65.22**%	54.35%
ResNet50	52.17%	60.87%
WideResNet101	60.87%	60.87%
SqueezeNet	56.52%	**63.04** %
EfficientNet	63.04%	58.70%

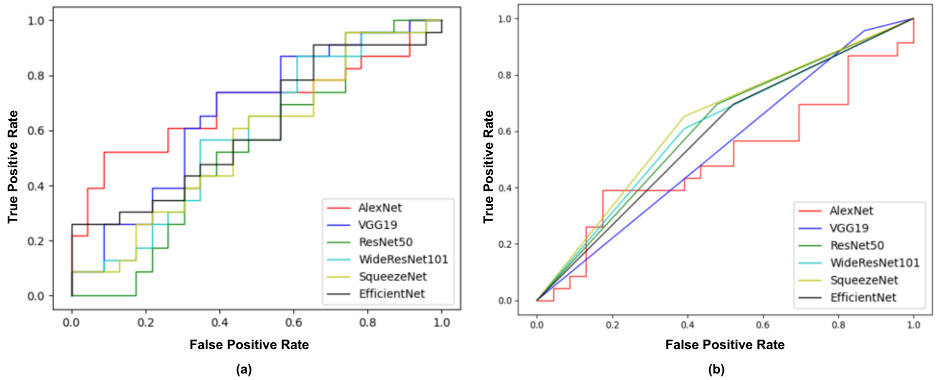

Figure 3. ROC curves of (a) the SVM, and (b) DL classifiers.

VGG model achieved the best classification results with obtained accuracy of 65.22% compared to the end-to-end deep learning classifier. Meanwhile, the deep learning classifier trained with radiomics from pre-trained SqueezeNet obtained a competitive classification accuracy of 60.04%. From the overall results in Table 2, we can say that for the transfer learning approach where deep learning models trained on non-medical data and are used to extract features from unseen DBT images, the SVM classifier has a promising classification accuracy and surpassed Deep learning classification on approximately all evaluated model.

Figure 3 shows the ROC curves for SVM and deep learning classifier for the transfer learning approach of all the evaluated pre-trained models. It is shown that the SVM classifier trained with VGG deep features (radiomics) performs best with an AUC of 0.65 compared to other feature extraction models. The second-best performing classifier is the deep learning classifier trained with wideResNet radiomics, even if its overall accuracy is lower than features from SqeezeNet with an AUC of 0.67.

Regarding class classification accuracy, Figure 4 presents the confusion matrix for the

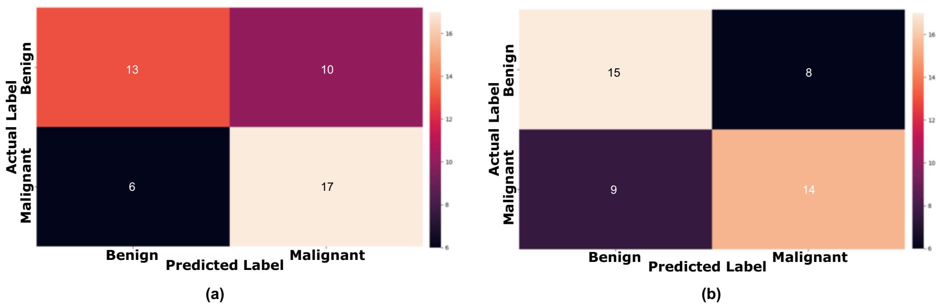

Figure 4. Confusion matrices of (a) the SVM classifier trained with deep features of VGG19 Network and (b) the DL classifier trained with deep features of SqueezeNet Network.

best SVM classifier with radiomics from the pre-trained VGG model and the best end-to-end deep learning classification with radiomics from the pre-trained SqueezeNet model. As one can see, the SVM classifier succeeded in classifying malignant tumors by more than 73%, outperforming the deep learning classifier by approximately 13%, while the deep learning classifier outperformed the SVM classifier in classifying benign tumors by 4%.

3.2. *Performance evaluation of fine-tuning approach*

In the case of the fine-tuning method, Table 3 compares the classification performance of the SVM classifier and the end-to-end deep learning classifier for the evaluated fine-tuned deep feature (Radiomics) extractor models trained with the training set of the DBT images dataset. As shown, when comparing values from Table 2 to values from Table3, we can argue that fine-tuning the deep learning-based radiomics can yield noticeable improvements in terms of classification accuracy for both classifiers, especially for the end-to-end deep learning classifier.

With the fine-tuning approach, the performance of the SVM classifier trained with radiomics from the AlexNet model increased by 8.7%. Besides, the classification accuracy of the end-to-end deep learning classifier of the AlexNet model was also advanced by 23.91%. It is also noticeable that there is a comparative performance in the results of the deep learning classifier for the rest of the models, with classification accuracy ranging from 65.22% to 73.91%.

Table 3. The performance of the SVM classifier and deep learning classifier with the fine-tuning approach.

Model	SVM classifier	Deep learning classifier
AlexNet	**71.74** %	**80.43**%
VGG19	60.87%	67.39%
ResNet50	52.17%	65.22%
WideResNet101	50.00%	73.91%
SqueezeNet	60.87%	69.57%
EfficientNet	60.87%	67.39%

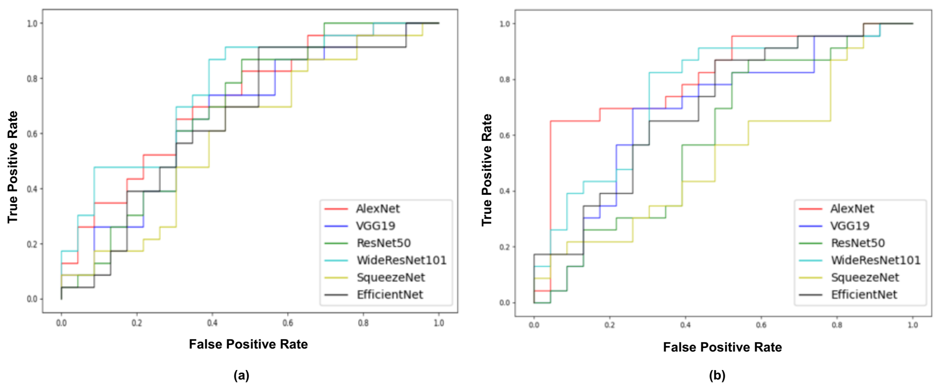

Figure 5. ROC curves of (a) the SVM, and (b) DL classifiers.

The ROC curves for SVM and deep learning classifier for the fine-tuning method are shown in Figure 5. It is shown that the SVM classifier and the deep learning classifier trained with radiomics from the AlexNet model have the best performance, confirming the high efficiency of both classifiers presented in Table 3 obtaining AUC of 0.72 and 0.80, respectively.

Regarding class classification accuracy, Figure 6 shows the confusion matrix for the best

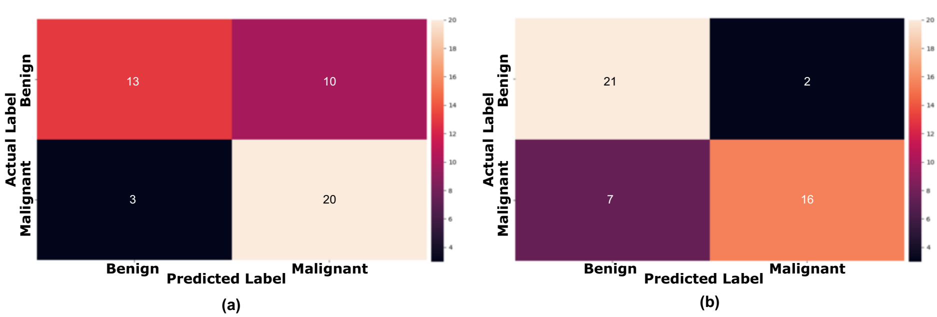

Figure 6. Confusion matrices of (a) the SVM classifier trained with deep features of AlexNet Network and (b) the DL classifier trained with deep features of AlexNet Network.

SVM classifier with radiomics from the fine-tuned AlexNet model and the best end-to-end deep learning classification with the fine-tuned AlexNet model. Here, the deep learning classifier has better accuracy in classifying the benign tumors with obtained accuracy of 91%, while the SVM classifier outperformed in classifying Malignant tumors by 17%. Based on the above analysis, we can conclude that to transfer learning techniques to extract features in cooperation with the SVM classifier is better for DBT images than the deep learning classifier. In contrast, end-to-end deep learning classification based on radiomics from fine-tuned models can significantly improve breast tumor classification accuracy. Of note, the classification results could be further improved by utilizing an optimization algorithm, such as the stochastic whale optimization algorithm [25] to find the optimal parameters of SVM.

4. Conclusions

This paper presents a breast tumor classification method for digital breast tomosynthesis images (DBT) based on radiomics extracted from the most innovative deep learning classification models. Our work first investigated the transfer learning technique where pre-trained on non-medical images CNN models are used to directly extract radiomics from the medical DBT images to train support vector machine and deep learning-based classifier. Secondly, the fine-tuning technique where the CNN models pre-trained on non-medical images is fine-tuned on the training set of DBT images. With the only publicly available digital breast tomosynthesis dataset, our experiments showed that end-to-end deep learning classification with radiomics extracted from the fine-tuned AlexNet model achieved the best classification accuracy of 80.43%.
The future work will focus on developing a breast tumor classification method based on the aggregation of robust deep learning-based feature extraction models.

Acknowledgement

The Spanish Government partly supported this research through Project PID2019-105789RB-I00.

References

[1] Gunjan Chugh, Shailender Kumar, and Nanhay Singh. Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation*, 13(6):1451–1470, jan 2021.

[2] Mohamed Abdel-Nasser, Jaime Melendez, Antonio Moreno, and Domenec Puig. The impact of pixel resolution, integration scale, preprocessing, and feature normalization on texture analysis for mass classification in mammograms. *International Journal of Optics*, 2016.

[3] Mohamed Abdel-Nasser, Antonio Moreno, and Domenec Puig. Temporal mammogram image registration using optimized curvilinear coordinates. *Computer methods and programs in biomedicine*, 127:1–14, 2016.

[4] Vivek Kumar Singh, Mohamed Abdel-Nasser, Farhan Akram, Hatem A Rashwan, Md Mostafa Kamal Sarker, Nidhi Pandey, Santiago Romani, and Domenec Puig. Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. *Expert Systems with Applications*, 162:113870, 2020.

[5] Mohamed Abdel-Nasser, Adel Saleh, Antonio Moreno, and Domenec Puig. Automatic nipple detection in breast thermograms. *Expert Systems with Applications*, 64:365–374, 2016.

[6] JAMES V. FIORICA. Breast cancer screening, mammography, and other modalities. *Clinical Obstetrics and Gynecology*, 59(4):688–709, dec 2016.

[7] Srinivasan Vedantham, Andrew Karellas, Gopal R. Vijayaraghavan, and Daniel B. Kopans. Digital breast tomosynthesis: State of the art. *Radiology*, 277(3):663–684, dec 2015.

[8] Xuan-Anh Phi, Alberto Tagliafico, Nehmat Houssami, Marcel J. W. Greuter, and Geertruida H. de Bock. Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts – a systematic review and meta-analysis. *BMC Cancer*, 18(1), April 2018.

[9] Kathryn P. Lowry, Rebecca Yates Coley, Diana L. Miglioretti, Karla Kerlikowske, Louise M. Henderson, Tracy Onega, Brian L. Sprague, Janie M. Lee, Sally Herschorn, Anna N. A. Tosteson, Garth Rauscher, and Christoph I. Lee. Screening performance of digital breast tomosynthesis vs digital mammography in community practice by patient age, screening round, and breast density. *JAMA Network Open*, 3(7):e2011792, July 2020.

[10] Mihalj Bakator and Dragica Radosav. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, 2(3):47, August 2018.

[11] Parita Oza, Paawan Sharma, Samir Patel, and Pankaj Kumar. Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey. *Neural Computing and Applications*, 34(3):1815–1836, jan 2022.

[12] Hiba Chougrad, Hamid Zouaki, and Omar Alheyane. Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, 157:19–30, apr 2018.

[13] Adeyinka P. Adedigba, Steve A. Adeshina, and Abiodun M. Aibinu. Performance evaluation of deep learning models on mammogram classification using small dataset. *Bioengineering*, 9(4):161, apr 2022.

[14] Inês C. Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S. Cardoso. INbreast. *Academic Radiology*, 19(2):236–248, feb 2012.

[15] Vitoantonio Bevilacqua, Antonio Brunetti, Andrea Guerriero, Gianpaolo Francesco Trotta, Michele Telegrafo, and Marco Moschetta. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognitive Systems Research*, 53:3–19, jan 2019.

[16] Ravi K. Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark A. Helvie, Caleb D. Richter, and Kenny H. Cha. Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Transactions on Medical Imaging*, 38(3):686–696, mar 2019.

[17] Yu Zhang, Xiaoqin Wang, Hunter Blanton, Gongbo Liang, Xin Xing, and Nathan Jacobs. 2d convolutional neural networks for 3d digital breast tomosynthesis classification. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, nov 2019.

[18] SPIE-AAPM-NCI DAIR Digital Breast Tomosynthesis Lesion Detection Challenge. https://www.aapm.org/GrandChallenge/DBTex/. Accessed: 2022-05-20.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84 – 90, 2012.

[20] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, November 2015.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2016.

[22] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Procedings of the British Machine Vision Conference 2016*. British Machine Vision Association, 2016.

[23] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5mb model size, 2016.

[24] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.

[25] FatmaAlzahra Mohamed, Mohamed Abdel-Nasser, Karar Mahmoud, and Salah Kamel. Economic dispatch using stochastic whale optimization algorithm. In *2018 International Conference on Innovative Trends in Computer Engineering (ITCE)*, pages 19–24. IEEE, 2018.