Breast Cancer Classification in Ultrasound Images using Transfer Learning

Ahmed Hijab
Biomedical Engineering and Systems
Faculty of Engineering
Cairo University, Egypt
engahmadhijab@gmail.com

Muhammad A. Rushdi
Biomedical Engineering and Systems
Faculty of Engineering
Cairo University, Egypt
mrushdi@eng1.cu.edu.eg

Mohammed M. Gomaa
Department of Diagnostic Radiology
Faculty of Medicine
Cairo University, Egypt
m.gomaa@gmail.com

Ayman Eldeib
Biomedical Engineering and Systems
Faculty of Engineering
Cairo University, Egypt
eldeib@ieee.org

Abstract— Computer-aided detection of malignant breast tumors in ultrasound images has been receiving growing attention. In this paper, we propose a deep learning methodology to tackle this problem. The training data, which contains several hundred images of benign and malignant cases, was used to train a deep convolutional neural network (CNN). Three training approaches are proposed: a baseline approach where the CNN architecture is trained from scratch, a transfer-learning approach where the pre-trained VGG16 CNN architecture is further trained with the ultrasound images, and a fine-tuned learning approach where the deep learning parameters are fine-tuned to overcome overfitting. The experimental results demonstrate that the fine-tuned model had the best performance (0.97 accuracy, 0.98 AUC), with pre-training on US images. Creating pre-trained models using medical imaging data would certainly improve deep learning outcomes in biomedical applications.

Keywords— Breast lesion; ultrasound; convolutional neural networks; deep learning; transfer learning.

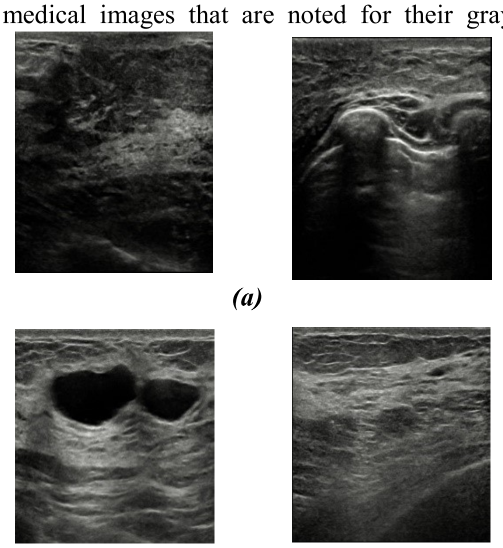
I. INTRODUCTION

Breast cancer is the most common cancer in women world-wide. The National Cancer Institute of the United States of America predicted the number of new breast cancer patients in 2018 to be 268,270 [1]. In addition, 38.8% of Egyptian women diagnosed with cancer, are breast cancer patients [2].

A suspicious result can be followed up by further diagnosis. However, mammography sometimes shows up a suspicious area that is not cancer which can lead to unnecessary stress and sometimes interventions. Moreover, an MRI scan involves injecting a dye into the patient, hence, find out how far the cancer has spread [5][12]. An ultrasound scan can help differentiate between a solid mass or a fluid-filled cyst. In an ultrasound breast image, some features help distinguish whether a breast tumor, nodule, or lump is benign or malignant as shown in Fig.1. It can be summed up that a benign ultrasound breast image has smooth highly regular appearance, looks brighter in general, and may have two or three smooth lobulations. On the other hand, a malignant ultrasound breast image is usually appearing with a shape of mass with more height than width, angular margins where the mass edges are hard to define, or, with an increased presence of blood vessels [6].

Deep convolutional neural networks (CNN) have become of great and popular use in machine learning and computer-aided detection applications [13]. The CNN capabilities can be managed by controlling their depth and breadth where the accuracy of assumptions is high when it comes to nature-based datasets. On the other hand, when it comes to medical imaging, its usefulness has been witnessed for classification, segmentation, and detection. However, some drawbacks are faced due to the lack of medical training data of large numbers of images, as well as, the absence of deep learning models pre-trained on medical data. To help overcome such drawbacks, transfer learning is used in two ways. First, by employing the pre-trained architecture as a feature extractor that will minimize the dimensionality of the dataset using it as an input to a narrow classifier, and second, through fine-tuning by which layers can be specified to freeze, other layers' learning can be enhanced, and modifying the output size of the end layer [17].

We assume that the small pool of domain-specific CNN architectures necessitates further research into CNN models trained from scratch, and CNN models trained on natural images. However, in previous work by Hadad *et al.* [17], medical images that are noted for their gray scale,



(a)

(b)

Figure 1: Samples of the ultrasound breast lesion images used in this paper (a) Malignant, (b) Benign.

texture, and low contrast, are unlike other natural images on which models like VGG-Net are trained, and therefore, are expected to have less accuracy with medical images. In their work, they used mammogram (MG) and MRI datasets of 282 and 123 images, respectively, for transfer learning on VGG-Net. For this reason, in this paper we plan to extend their research and experiment on 1,000 ultrasound images instead, that were further augmented, using a CNN trained from scratch, VGG transfer learning, and, fine-tuned transfer-learning CNN.

II. RELATED WORK

Several papers have been written on breast cancer detection and many approaches have been proposed. We review briefly deep learning approaches: transfer learning, FCN-AlexNet, patch-based LeNet, U-Net, Nakagami parametric imaging with CNN, and fine-tuned VGG-Net.

Yap *et al.* [18] constructed two ultrasound image datasets (A and B) and showed that a transfer-learning FCN-AlexNet method had the highest accuracy (0.92 for Dataset A, 0.88 for Dataset B). Results on other architectures were comparable: patch-based LeNet (0.91 for Datasets A and B), and the U-Net method (0.89 for Dataset A, 0.78 for Dataset B). Their work was compared against deformable part models (DPM) and other deep learning approaches. However, Yap *et al.* communicated the need of such approaches for an extensive training process and negative images.

Byra *et al.* [16] employed Nakagami parameter maps to train a CNN with an area under the receiver operating characteristic curve (AUC) equivalent to 0.91. In their work, they classified 458 RF ultrasound images. They concluded that the Nakagami method can be affected by quantitative ultrasound parameters (QUS), and they also confirmed that the CNN classification is correlated directly with the image quality.

While Hepsag *et al.* [19] employed CNN on 2 sets of data: mini-MIAS [21], and BCDR [19]. For baseline experiments, they got performance measures between 60-72%. When they used preprocessing operations (such as cropping, and data augmentation), the accuracy on the BCDR dataset increased from 65 to 85%. Similar enhancements were obtained on the mini-MIAS dataset.

Abdel-Zaher and Eldeib [20] used a deep belief network (DBN) to classify breast cytology images. They achieved an accuracy of 99.68%, a 100% sensitivity and a 99.47% specificity. They employed DBN in unsupervised manner to recognize the characteristics of the Wisconsin Breast Cancer Dataset (WBCD). Further, a supervised phase was made based on a back-propagation deep architecture which exploits the conjugate gradient and the Levenberg-Marquardt optimization algorithms. The authors confirmed that the accuracy of their proposed network model (DBN-NN) is better than that of the randomly initialized weight backward propagation NN (RIW-BPNN) for both of the conjugate gradient and the Levenberg-Marquardt algorithms.

Hadad *et al.* [17] investigated three CNN models for MRI breast lesion classification, namely de novo training, cross-domain transfer learning using VGG-Net, and cross-modal transfer learning using mammography images. The authors showed accuracies of 0.94, 0.90, and 0.93, respectively. The authors used augmented datasets of mammogram images (282 augmented to 32,064) and MRI images (123 augmented to 19,316). First, they employed a CNN trained from scratch by MR images. Second, they tried a fine-tuned CNN through VGG-Net. Finally, they created a fine-tuned CNN through MG-Net.

In our work, we focus on breast lesion image classification using three variants of deep architectures: a CNN architecture trained from scratch with an augmented ultrasound dataset, a transfer learning VGG16 architecture further trained with US images, and a fine-tuned architecture where the deep learning parameters are fine-tuned for better performance.

III. MATRIALS & METHODOLOGY

In this work, the data was collected at Baheya Foundation for Treatment of Breast Cancer in Egypt. Ultrasound images were produced by the GE Ultrasound LOGIQ E9 XDclear model with a broad-spectrum linear matrix array probe (ML6-15-D). Data collection was carried on by expert doctors who marked the lesions and their margins according to the BI-RADS scoring criteria with approvals of the Baheya Ethics Committee and the Ministry of Health and Population in Egypt.

The classification model in this paper was built using the Keras library, which is a simple and efficient deep learning library. A small network was trained from scratch (as a baseline model) where Python version 3.5 with Theano backend was used in classifying tumors. The training was run with 1000 images, 50 epochs, a batch size of 20, and a learning rate of 0.001. Two more network models were used. The first model, VGG16, is a pre-trained deep convolutional neural network (CNN). VGG16 was trained using more than 14 million images of 1000 classes. The second model is a fine-tuned model derived from the saved features of the VGG16 pre-trained CNN model. The main building blocks in CNN are rectified linear units (ReLU), pooling, dropout layers, connected layers, and loss layers [14].

A. Dataset

In our work, we collected a dataset that consists of 1300 high-resolution ultrasound images in JPEG format, with a size of 960×720 pixels for each image. The dataset was divided into a 1,000-image training set (650 benign and 350 malignant), and a 300-image test set (165 benign and 135 malignant).

The first step in our pipeline is to enlarge the dataset and reduce class imbalance through rotation (90º, 180º, 270º), width shift, height shift, rescaling, shearing, zooming, horizontal flipping, and filling. The importance of this step lies in its ability to increase the size of our data by 24 times to reach 21,600 images, and therefore, improve the generalization performance. The augmented dataset was divided into 2 portions: 70% for training, and 30% for testing. This augmentation process can help overcome overfitting, which is also reduced by the drop-out layers.

B. *Pre-trained model*

VGG16 is a 16-layer deep learning model that has been trained to classify images into 1000 categories [11]. Models initialized with the VGG16 model do not need big numbers of labeled data or excessive computations. Hence, the VGG16 pre-trained model has been used to accelerate the training of deep models for other problems through transfer learning [10,13].

C. *Pre-trained fine-tuned model*

The parameters of the VGG16 model are used as an initialization of a fine-tuned model for the dataset under consideration. All of the convolutional layers were frozen except for the last one. Moreover, the stochastic gradient descent (SGD) algorithm is used to update the network weights using the breast tumor dataset [15]. Classifier training was carried on with 50 iterations.

D. *Experiments*

The experiments were carried out using NVIDIA GeForce GTX 1050 GPU where the baseline model required a duration of 5 minutes (for model training and testing), the VGG16 model required the duration of 100 second (for model testing), and the fine-tuned method required 36.6 minutes (for training and testing).

IV. Results

Figure 2 shows the temporal improvement in the accuracy of the detection of benign and malignant breast lesions by the baseline method where the deep architecture is trained from scratch.

Results of the second approach based on transfer learning and the pre-trained VGG16 model are shown in Figure 3. Finally, results after further fine-tuning of the deep learning parameters are shown in Figure 4. Figures 2-4 show the relative performance among the three approaches.

Eventually, the fine-tuned learning approach using US images outperforms the transfer-learning and baseline approaches, as well as, the MG and MR fine-tuned learning approaches [17].

Prior to fine-tuning, as shown in Table 1, an MR-based breast cancer detection model trained from scratch has an accuracy of 0.94 and AUC of 0.98. This performance is better than that of US-based detector trained from scratch, which shows an accuracy of 0.79 and an AUC of 0.81. The US-based model constructed using the pre-trained VGG16 model has an accuracy of 0.95 and an AUC of 0.97. This performance is similar to that of the MR-based detector trained from scratch. However, after fine-tuning, the accuracy of the US-based fine-tuned model is the highest: 0.973 and the AUC is 0.98. This outperforms the mammogram-based fine-tuned detector which has a 0.93 accuracy and a 0.97 AUC.

Table 1 Comparison between accuracy of our tested model using US images and other related work.

Task	*Accuracy*	*AUC*
Hadad et al. MR trained from scratch	*0.94*	*0.98*
Hadad et al. MR VGG-Net	*0.90*	*0.95*
Hadad et al. fine-tuned MG-Net	*0.93*	*0.97*
Our tested – US Baseline method	*0.79*	*0.81*
Our tested – US Pre-trained VGG-Net	*0.9511*	*0.97*
Our tested – US Fine-tuned VGG	*0.9739*	*0.98*

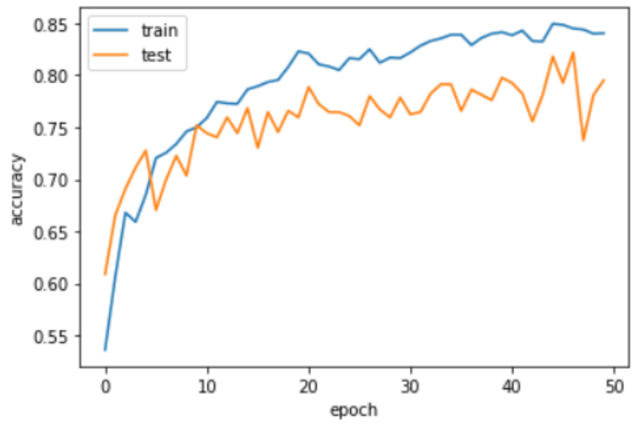


Figure 2: Detection accuracy using the baseline method.

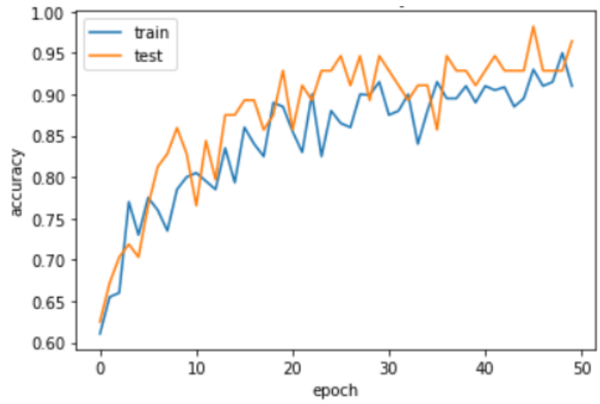


Figure 3: The system accuracy using transfer learning and the pre-trained VGG16 model.

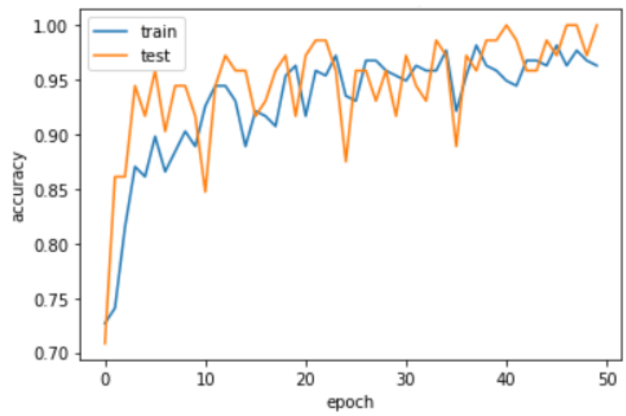


Figure 4: The system accuracy using the VGG16 pre-trained model and the fine-tuning method.

V. Discussion and Conclusions

We investigated three variants of a deep learning approach for computer-aided detection of malignant breast tumors in ultrasound images. We have shown experimentally that the fine-tuned variant of the pre-trained VGG16 model has the best performance metrics.

In a nutshell, the results have an accuracy range from 79% to 97%. To sum up our steps, we used data augmentation to increase the number of training images, and overcome the overfitting by width shift, height shift and rotation. Also, we used the bottleneck features of the VGG16 pre-trained model and applied fine-tuning to increase accuracy.

The results assert that the fine-tuned model with medical data for pre-training has improved the accuracy of classification, although our dataset is relatively small. This work represents a step towards the creation of practical and adaptable deep learning systems for breast cancer classification. These systems should supplement and support the clinical assessment and treatment approaches [3, 4, 7]. Although state-of-the-art ultrasound-based methods have been developed towards achieving high-performance breast cancer detection [22, 23], several challenges are still largely open. In particular, the lack of demographic variations of race and ethnicity in the training data can negatively impact the detection and survival outcomes for underrepresented patient groups [24]. Our work attempts to fill this gap through exploiting ultrasound breast images collected from North African and Middle Eastern patients. For future work, we seek to create a deep learning architecture with pre-training data collected from different imaging modalities. This pre-trained model can be useful for devising new automated detection systems based on medical imaging.

VI. Acknowledgments

This work was supported by Baheya Foundation for Treatment of Breast Cancer in Egypt and its department of diagnostic radiology.

References

[1] R. A. Smith, D. Brooks, V. Cokkinides, D. Saslow, and O. W. Brawley, “Cancer screening in the United States, 2013,” *CA: A Cancer Journal for Clinicians*, vol. 63, no. 2, pp. 87–105, 2013.

[2] A. S. Ibrahim, H. M. Khaled, N. N. Mikhail, H. Baraka, and H. Kamel, “Cancer Incidence in Egypt: Results of the National Population-Based Cancer Registry Program,” *Journal of Cancer Epidemiology*, vol. 2014, pp. 1–18, 2014.

[3] C. Vogel, "Efficacy and Safety of Trastuzumab as a Single Agent in First-Line Treatment of HER2-Overexpressing Metastatic Breast Cancer", Journal of Clinical Oncology, vol. 20, no. 3, pp. 719-726, 2002.

[4] I. Mayer, R. Dent, T. Tan, P. Savas and S. Loi, "Novel Targeted Agents and Immunotherapy in Breast Cancer", American Society of Clinical Oncology Educational Book, vol. 37, pp. 65-75, 2017.

[5] S. Adams, P. Schmid, H. S. Rugo, E. P. Winer, D. Loirat, A. Awada, D. W. Cescon, H. Iwata, M. Campone, R. Nanda, R. Hui, G. Curigliano, D. Toppmeyer, J. Oshaughnessy, S. Loi, S. Paluch-Shimon, D. Card, J. Zhao, V. Karantza, and J. Cortes, “Phase 2 study of pembrolizumab (pembro) monotherapy for previously treated metastatic triple-negative breast cancer (mTNBC): KEYNOTE-086 cohort A.,” *Journal of Clinical Oncology*, vol. 35, no. 15_suppl, pp. 1008–1008, 2017.

[6] M. Gnant, C. V. Poznak, and L. Schnipper, “Therapeutic Bone-Modifying Agents in the Nonmetastatic Breast Cancer Setting: The Controversy and a Value Assessment,” *American Society of Clinical Oncology Educational Book*, vol. 37, pp. 116–122, 2017.

[7] H. W. Hazard and N. M. Hansen, “Image-Guided Procedures for Breast Masses,*” Advances in Surgery*, vol. 41, pp. 257–272, 2007.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever & R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *J. Machine Learning Res*, vol. 15, pp.1929–1958, 2014.

[9] R. Girshick, “Fast R-CNN,” *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] K.Simonyan, A.Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[12] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, “Deep Learning for Health Informatics,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.

[13] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M. Lew, "Deep learning for visual understanding: A review", Neurocomputing, vol. 187, pp. 27-48, 2016.

[14] J.D. Gallego - Posado, D.A. Montoya-Zapata, and Q.L. Quintero-Montoya, “Detection and diagnosis of breast tumors using deep Convolutional Neural Networks,” Research Group on Mathematical Modeling School of Mathematical Sciences Universidad EAFIT Medell´Õn, Colombia.

[15] Z. You and B. Xu, "Improving training time of deep neural networkwith asynchronous averaged stochastic gradient descent," *The 9th International Symposium on Chinese Spoken Language Processing*, Singapore, 2014, pp. 446-449. doi: 10.1109/ISCSLP.2014.6936596

[16] M. Byra, H. Piotrzkowska-Wróblewska, K. Dobruch-Sobczak and A. Nowicki, "Combining Nakagami imaging and convolutional neural network for breast lesion classification," 2017 IEEE International Ultrasonics Symposium (IUS), Washington, DC, 2017, pp. 1-4.

[17] O. Hadad, R. Bakalo, R. Ben-Ari, S. Hashoul and G. Amit, "Classification of breast lesions using cross-modal deep learning," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 109-112.

[18] M. H. Yap et al., "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1218-1226, July 2018.

[19] P. U. Hepsağ, S. A. Özel and A. Yazıcı, "Using deep learning for mammography classification," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 418-423.

[20] A. Abdel-Zaher and A. Eldeib, "Breast cancer classification using deep belief networks", Expert Systems with Applications, vol. 46, pp. 139-144, 2016.

[21] K L. Kashyap, M. K. Bajpai and P. Khanna, "Breast cancer detection in digital mammograms," 2015 IEEE International Conference on Imaging Systems and Techniques (IST), Macau, 2015, pp. 1-6. doi: 10.1109/IST.2015.7294523

[22] M. H. Yap *et al*., "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218-1226, July 2018.

[23] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim and K. M. Lee, "Joint Weakly and Semi-Supervised Deep Learning for Localization and Classification of Masses in Breast Ultrasound Images," in *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 762-774, March 2019.

[24] B. Daly, and O. I. Olopade, "Race, ethnicity, and the diagnosis of breast cancer," in *JAMA*, vol. 313, no. 2, pp. 141–142, 2015.