This paper was a finalist for the Robert F Wagner F. All Conference Best Student Paper Award

# Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning

Bahareh Behboodi[a], Hamze Rasaee[a], Ali K. Z. Tehrani[a], and Hassan Rivaz[a,b]

[a]Electrical and Computer Eng., Concordia University, Montreal, Canada
[b]PERFORM center, Concordia University, Montreal, Canada

## ABSTRACT

Ultrasound (US) is a low-cost, portable, and safe tool for breast cancer screening. However, automatic classification of invasive ductal carcinoma (IDC) in US is a difficult classification task due to their similar appearance to fibroadenoma (FA) (a type of benign tumor). Another challenge is the limited availability of US data with ground truth labels, further complicating the adoption of deep learning techniques for IDC detection. It has been shown that deep classification networks perform better when they simultaneously learn multiple correlated tasks. However, most previous studies on breast US classifications focused on the binary classification of benign versus malignant tumors. To this end, we propose a multi-class classification deep learning-based strategy mainly focusing on the classification of IDC. Inspired by multi-task learning (MTL), we adopt a novel scheme in adding the background tissue as an additional class and show substantial improvements in IDC detection.

**Keywords:** Classification, Breast Lesion, Ultrasound.

## 1. INTRODUCTION

Breast cancer is one of the most common leading causes of the death among women worldwide. For screening the breast cancers, mammography is usually used as the primary imaging modality. However, due to low specificity of mammography in distinguishing dense and cancerous tissues, and the use of harmful radiation, ultrasound (US) has been used as an important alternative for breast cancer screening.[1] US as a portable, cost-effective, and non-invasive diagnosis modality utilizes low energy US waves and provides real-time examinations. In previous methods of breast US classification, US images were either used in full size or divided into subregions, and then texture related features were manually extracted and input into a classifier (i.e. support vector machine (SVM), random forest (RF), etc).[2–5] The main focus of previous works was the extraction of hand-engineered features, more precisely, texture-related features from a US image.

Recent state-of-the-art deep learning methods have paved the way for automatically extracting the most meaningful features by adopting convolution layers. The promising results of deep learning methods in the domain of medical images for classification tasks have raised researchers' attentions.[6–10] Ting *et al.* adopted a prior detection step before classifying breast lesions in mammography images.[8] In US breast classification, Han *et al.*,[6] employed a pre-trained GoogleNet with some modifications for classifying benign and malignant lesions of histogram equalized of 7408 US images. In most of the studies related to classification of US breast lesions, the main focus is the binary classification of benign versus malignant lesions.[11–13] Consequently, the adopted deep learning network learns only one task. In the domain of multi-task learning (MTL) studies, it has been shown that networks perform better when they are assigned with multiple tasks compared to only one task.[14]

Due to appearance of invasive ductal carcinomas, the task of distinguishing them from fibroadenoma is the most challenging task compared to binary classification of benign versus malignant. Therefore, we propose a deep learning-based scheme that performs better when it is assigned to multi-class classification task. To be more specific, we propose multi-class classification of fibroadenoma, cyst, and invasive ductal carcinomas in US images with limited data. We further propose a novel technique in taking the background of US image into account as an additional class leading to a 4-class classification task for breast US images. We also show that the proposed scheme of multi-class classification including background as the additional class, holds for different deep learning networks. To cope with uncertainty in network's estimations, we adopt test-time augmentation for classification evaluation.[15] Our contributions can be summarized as below:

- Multi-class classification of breast US with limited available images
- A novel technique in adding background as an additional class
- Our proposed scheme holds for different networks
- Test-time augmentation for evaluation

## 2. METHOD

### 2.1 Dataset

In this study we used a publicly available US dataset.[16] It was collected in 2012 from the UDIAT Diagnostic Centre of the Parc Taul Corporation, Sabadell (Spain) with a 17L5 HD linear array transducer (8.5 MHz). The dataset consisted of a total of 163 images with a mean image size of 760 × 570 pixels including one or more lesions. Out of 163 lesions, 53 images had cancerous masses, which included the subcategories of 40 invasive ductal carcinomas, 4 were ductal carcinomas *in situ*, 2 invasive lobular carcinomas, and 7 unspecified malignant lesions. The remaining 110 were benign lesions (65 cysts, 39 fibroadenomas, and 6 another type of benign lesions). In this study, only 40 invasive ductal carcinomas (IDC), 65 cysts (CYST), and 39 fibroadenomas (FA) were used as the other classes contained very few samples. Figure 1 presents an example of US images used from the dataset.

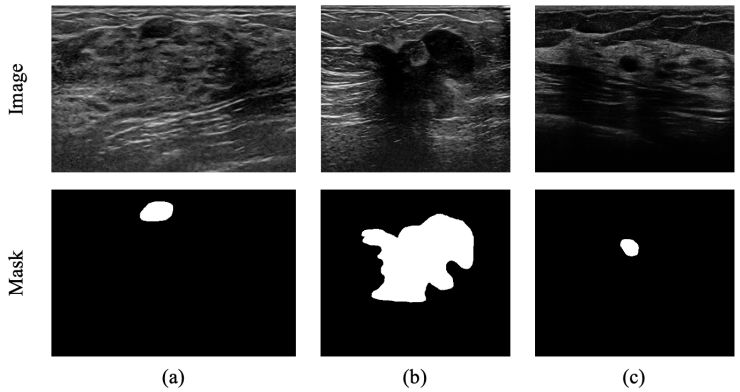


Figure 1: An example of breast US images from the dataset (a) FA, (b) IDC, and (c) CYST.

### 2.2 Preprocessing

As mentioned earlier, we propose a novel technique in adding background as an additional class. To this end, before feeding the images to the deep learning network, they were cropped in order to help the network learn the characteristics of each lesion type more precisely. Based on available segmentation masks of images, a marginal cropping window surrounding the lesion (blue window in Fig. 1) was used which was 40% larger than the window surrounding the lesion borders (red window in Fig. 1). All the cropped images were then resized to the size of 400 × 400, and their intensities were normalized to the range of 0 to 1. In order to keep balance between the number of lesion types in training and test sets, 80% of each lesion type was randomly selected for the training set and the rest was used in the test set. As for background class, a window was used to randomly crop the background (BG) of US image excluding any part of lesion. In US imaging, the more the US waves travel deeper, the more they are attenuated. Therefore, in order to have similar range of attenuation in BG class, the BG images were cropped from the same depth of lesion's location. In Fig. 1 the yellow square represents the area where BG images were cropped from. All the randomly selected BG images from the yellow area of Fig. 1 had the same size of the marginal cropping window (i.e. the blue window). In order to have balanced number of BG class, we randomly selected the BG images derived from all US images. As a result, in our training and validation sets the total number of FA, CYCT, IDC, and BG respectively were 31, 52, 32, and 38. Consequently, in our test set the number of each classes were 8, 13, 8, and 10, respectively.

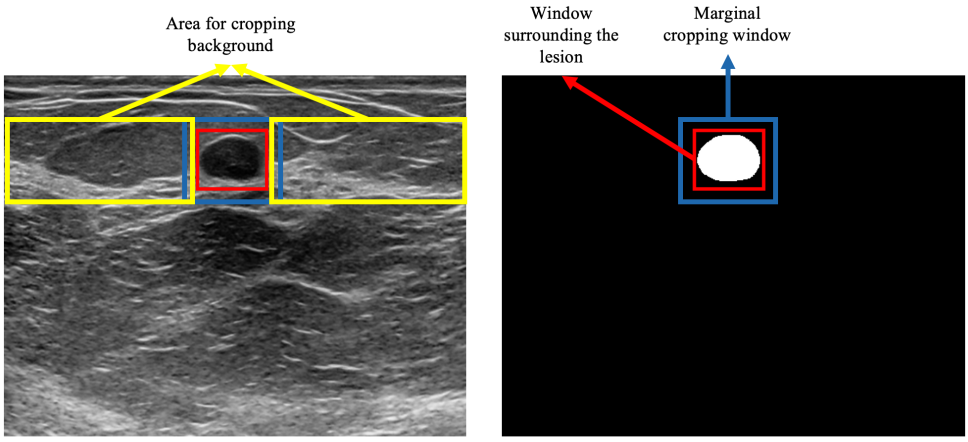


Figure 2: Cropping schematic used in this study. The red window is based on the lesion border in the mask. The blue window is 40% larger than red window. The yellow rectangles show the desire area for selecting BG images.

### 2.3 Experiments

Due to the small number of training data and large inter-class variations in the database, training the network from scratch was impossible. Therefore, we used two pre-trained networks, ResNet-34 and MobileNet-v2, separately as backbone feature extractors with some modifications for our dataset. A schematic of our network is shown in Fig. 3. In order to show the impact of multi-task learning on the performance of our network, we present our results based on two avenues: *2-class Avenue* and *4-class Avenue*. Our *2-class Avenue* was a 2-class classification problem (IDC versus all other classes), versus *4-class Avenue* was a 4-class classification (FA versus Cyst versus IDC versus BG). In both avenues, a random on-the-fly data augmentation of horizontal flip, width and height shifts, and zooming, was applied to the batches during training. The number of batches were set to 25 and training lasted for 100 epochs while saving the best model based on the validation accuracy. Adam optimizer was used[17] and its learning rate was tuned using cyclical learning rate.[18] It worth noting that to mitigate the effect of imbalanced data during training, we used weighted cross-entropy loss function in both avenues wherein the weights were initialized based on the distribution of images in each class. For improving the predictions, we employed test-time augmentation where the same augmentation strategy in the training set, was applied to test set. Therefore, we enlarged the number of images in test set by augmenting each image 4 times leading to 195 (i.e. 39×5) images in the test set.

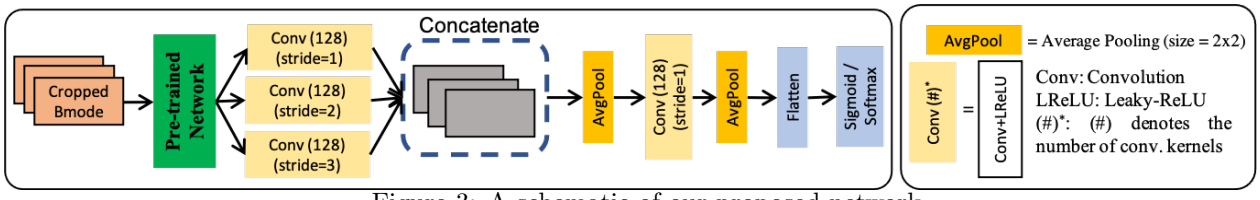


Figure 3: A schematic of our proposed network

## 3. RESULTS

The predicted labels were evaluated using receiver operating characteristic curve (ROC) and the area under curve (AUC) for both ResNet-34 and MobileNet-v2 in two proposed avenues. For ResNet-34, in *2-class Avenue* we achieved AUC of 0.66 for IDC. The AUC of IDC was improved by 31% in our *4-class Avenue* for the same network. Furthermore, the AUC scores for FA, Cyst, IDC, and BG in *4-class Avenue* were 0.87, 1.0, 0.87, and 1.0, respectively, for ResNet-34. Similarly, for MobileNet-v2, the AUC of IDC was improved from 0.82 in *2-class Avenue* to 0.90 in *4-class Avenue*. The AUC scores of FA, CYST, and BG for MobileNet-v2 were 0.87, 1.0, and 1.0, respectively.

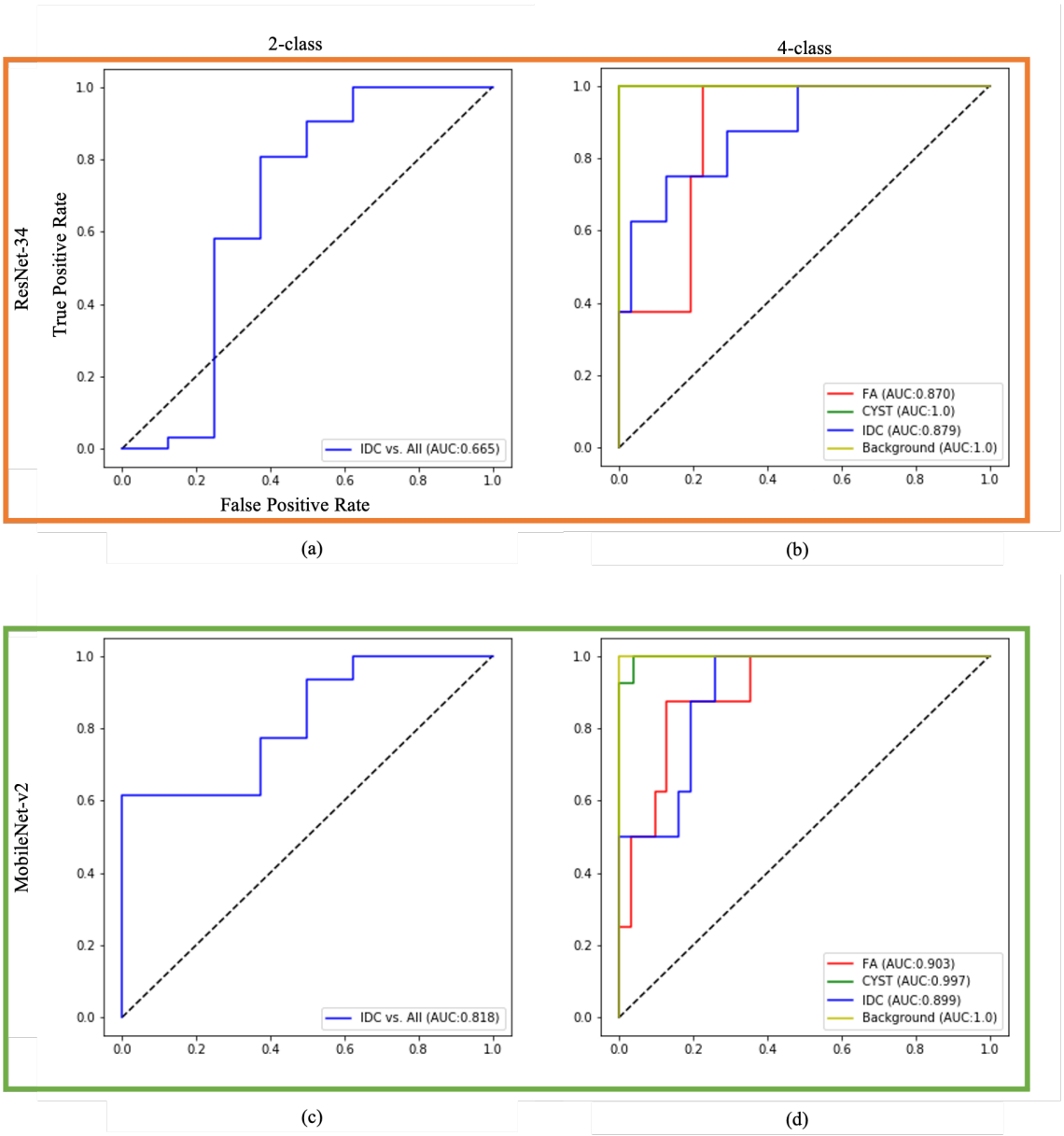


Figure 4: ROC curves and AUC results. The first and second rows show the ROC curves for ResNet-34 and MobileNet-v2, respectively. First column ((a) and (c)) presents the results for *2-class Avenue* whereas second column ((b) and (d)) present for *4-class Avenue*.

Table 1 summarizes the classification reports for IDC in two proposed avenues. For ResNet-34, the accuracy was improved by 18% in *4-class Avenue* showing that increasing the number of classes helped the network for better predictions. Precision and f1-score were improved from 0.38 to to 0.67 and 0.48 to 0.57, respectively. However, recall was decreased from 0.62 to 0.5 when using more classes (i.e. *4-class Avenue)*. Similarly, for MobileNet-v2, the accuracy was enhanced from 0.84 in *2-class Avenue* to 0.90 in *4-class Avenue*. We observed improvements in recall and f1-score for MobileNet-v2, however, precision dropped from 1.0 in *2-class Avenue* to 0.80 in *4-class Avenue*.

Table 1: Classification report for IDC.

| *Avenue* | *Accuracy* | *Precision* | *Recall* | *f1-score* |
|---|---|---|---|---|
| | | ***ResNet-34*** | | |
| 2-class | 0.71 | 0.38 | 0.62 | 0.48 |
| 4-class | 0.84 | 0.67 | 0.50 | 0.57 |
| | | ***MobileNet-v2*** | | |
| 2-class | 0.84 | 1.0 | 0.25 | 0.4 |
| 4-class | 0.90 | 0.8 | 0.50 | 0.62 |

## 4. CONCLUSION

Most of previous studies on US images classifications focused on binary classification of benign versus malignant lesions. However, the main challenge is in detection of FA versus IDC due to their similar appearance. In this study, we showed the importance of multi-task learning in better detection of IDC in breast US images. We investigated that increasing the number of classes led to better performance of the deep learning networks. We further proposed a novel strategy in adding background of US images as an additional class. We showed that our proposed scheme holds for different deap learning networks by adopting 2 pre-trained networks, ResNet-34 and MobileNet-v2. By adding more classes, we illustrated that the AUC score was improved by the factor of 31% and 9% for ResNet-34 and MobileNet-vs, respectively. Also, to control the network's uncertainty in its predictions, we adopted test-time augmentation.[19]

## 5. ACKNOWLEDGEMENT

This research was funded by Richard and Edith Strauss Foundation and by Natural Sciences and Engineering Research Council of Canada (NSERC) RGPIN-2020-04612.

## REFERENCES

[1] Cheng, H.-D., Shan, J., Ju, W., Guo, Y., and Zhang, L., "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern recognition* **43**(1), 299–317 (2010).

[2] Yang, M.-C., Moon, W. K., Wang, Y.-C. F., Bae, M. S., Huang, C.-S., Chen, J.-H., and Chang, R.-F., "Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis," *IEEE Transactions on Medical Imaging* **32**(12), 2262–2273 (2013).

[3] Gómez, W., Pereira, W. C. A., and Infantosi, A. F. C., "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE transactions on medical imaging* **31**(10), 1889–1899 (2012).

[4] Ding, J., Cheng, H.-D., Huang, J., Liu, J., and Zhang, Y., "Breast ultrasound image classification based on multiple-instance learning," *Journal of digital imaging* **25**(5), 620–627 (2012).

[5] Uniyal, N., Eskandari, H., Abolmaesumi, P., Sojoudi, S., Gordon, P., Warren, L., Rohling, R. N., Salcudean, S. E., and Moradi, M., "Ultrasound rf time series for classification of breast lesions," *IEEE transactions on medical imaging* **34**(2), 652–661 (2014).

[6] Han, S., Kang, H.-K., Jeong, J.-Y., Park, M.-H., Kim, W., Bang, W.-C., and Seong, Y.-K., "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Physics in Medicine & Biology* **62**(19), 7714 (2017).

[7] Becker, A. S., Mueller, M., Stoffel, E., Marcon, M., Ghafoor, S., and Boss, A., "Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study," *The British journal of radiology* **91**(xxxx), 20170576 (2018).

[8] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., "Dermatologist-level classification of skin cancer with deep neural networks," *nature* **542**(7639), 115–118 (2017).

[9] Ting, F. F., Tan, Y. J., and Sim, K. S., "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications* **120**, 103–115 (2019).

[10] Tehrani, A. K., Amiri, M., Rosado-Mendez, I. M., Hall, T. J., and Rivaz, H., "A pilot study on scatterer density classification of ultrasound images using deep neural networks," in [*2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*], 2059–2062, IEEE (2020).

[11] Byra, M., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., and Andre, M., "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion," *Medical physics* **46**(2), 746–755 (2019).

[12] Shin, S. Y., Lee, S., Yun, I. D., Kim, S. M., and Lee, K. M., "Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images," *IEEE transactions on medical imaging* **38**(3), 762–774 (2018).

[13] Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y., Lv, Q., and Yi, Z., "Automated diagnosis of breast ultrasonography images using deep neural networks," *Medical image analysis* **52**, 185–198 (2019).

[14] Ruder, S., "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098* (2017).

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

[16] Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A. K., and Martí, R., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics* **22**(4), 1218–1226 (2017).

[17] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[18] Smith, L. N., "Cyclical learning rates for training neural networks," in [*2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*], 464–472, IEEE (2017).

[19] Amiri, M., Brooks, R., Behboodi, B., and Rivaz, H., "Two-stage ultrasound image segmentation using u-net and test time augmentation," *International journal of computer assisted radiology and surgery* (2020).