End-to-end breast ultrasound lesions recognition with a deep learning approach

Moi Hoon Yap, Manu Goyal, Fatima Osman, Ezak Ahmad, Robert Martí, et al.

Moi Hoon Yap, Manu Goyal, Fatima Osman, Ezak Ahmad, Robert Martí, Erika Denton, Arne Juette, Reyer Zwiggelaar, "End-to-end breast ultrasound lesions recognition with a deep learning approach," Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, 1057819 (12 March 2018); doi: 10.1117/12.2293498

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

End-to-End Breast Ultrasound Lesions Recognition with a Deep Learning Approach

Moi Hoon Yap[a], Manu Goyal[a], Fatima Osman[b], Ezak Ahmad[a], Robert Martí[c], Erika Denton[d], Arne Juette[d], and Reyer Zwiggelaar[e]

[a]Manchester Metropolitan University, Chester Street, Manchester, UK
[b]Sudan University of Science and Technology, Khartoum, Sudan
[c]University of Girona, Girona, Spain
[d]Norfolk and Norwich University Hospital Foundation Trust, Norwich, UK
[e]Aberystwyth University, Aberystwyth, UK

ABSTRACT

Existing methods for automated breast ultrasound lesions detection and recognition tend to be based on multi-stage processing, such as preprocessing, filtering/denoising, segmentation and classification. The performance of these processes is dependent on the prior stages. To improve the current state of the art, we have proposed an end-to-end breast ultrasound lesions detection and recognition using a deep learning approach. We implemented a popular semantic segmentation framework, i.e. Fully Convolutional Network (FCN-AlexNet) for our experiment. To overcome data deficiency, we used a pre-trained model based on ImageNet and transfer learning. We validated our results on two datasets, which consist of a total of 113 malignant and 356 benign lesions. We assessed the performance of the model using the following split: 70% for training data, 10% for validation data, and 20% testing data. The results show that our proposed method performed better on benign lesions, with a *Dice* score of 0.6879, when compared to the malignant lesions with a *Dice* score of 0.5525. When considering the number of images with *Dice* score > 0.5, 79% of the benign lesions were successfully segmented and correctly recognised, while 65% of the malignant lesions were successfully segmented and correctly recognised. This paper provides the first end-to-end solution for breast ultrasound lesion recognition. The future challenges for the proposed approaches are to obtain additional datasets and customize the deep learning framework to improve the accuracy of this method.

Keywords: breast ultrasound lesions, breast cancer detection, fully convolutional network, AlexNet

1. INTRODUCTION

According to Breast Cancer Care [1], breast cancer is the most common cancer in the UK. One in eight women will be diagnosed with breast cancer in their lifetime and one person is diagnosed every 10 minutes [1]. Over recent years, there has been significant research into using different image modalities [2] and technical methods have been developed [3, 4] to aid early detection and diagnosis of the disease. These efforts have led to further research challenge and demand for robust computerised methods for cancer detection.

Two view mammography is known as the gold standard for breast cancer diagnosis [2]. However, ultrasound is the standard complementary modality to increase the accuracy of diagnosis. Other alternatives include tomography and magnetic resonance, however, ultrasound is the most widely available option and widely used in clinical practice [5].

Conventional computerised methods in breast ultrasound cancer diagnosis comprised multiple stages, including pre-processing, detection of the region of interest (ROI), segmentation and classification [6–8]. These processes rely on hand-crafted features including descriptions in the spatial domain (texture information, shape and edge descriptors) and frequency domain. With the advancement of deep learning methods, we can detect

Further author information: (Send correspondence to M.H. Yap)
M.H.Yap.: E-mail: m.yap@mmu.ac.uk, Telephone: +44 161 247 1503

Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Barjor Gimi, Andrzej Krol, Proc. of SPIE Vol. 10578, 1057819 · © 2018 SPIE
CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293498

and recognise objects without the need for hand-crafted features. This paper presents the limitation of the state of the art and conducts a feasibility study on the use of a deep learning approach as an end-to-end solution for fully automated breast lesion recognition in ultrasound images.

2. RELATED WORK

This section summarises the machine learning and deep learning methods for breast ultrasound cancer analysis. Most of the existing publications on breast ultrasound classification have focused on lesion classification, i.e. classify lesion as malignant or benign.

The majority of state-of-the-art methods are multi-stage. First to detect a lesion, i.e. where a lesion is localised on the image [9]. The localisation of a lesion can be done by manual annotation or using automated lesion detection approaches [6, 10]. Then to identify the lesion type using feature descriptors. Among different proposed approaches considering solid mass classification. There are two main feature descriptors [11], i.e. echo texture [12] [13] and shape and margin features [14]. We present a few multi-stage machine learning methods below, but for a full review, please refer to Cheng et al. [9].

In 2002, a novel computer-aided diagnosis (CADx) system using neural networks was proposed by Chen et al. [15]. First the region-of-interest (ROI) was located by a scanning physician. At the segmentation stage, they proposed an image processing algorithm to segment the lesion based on rule-based contour boundary approximation. At the feature extraction stage, three feasible features were extracted from the segmented ROI images, namely variance contrast, autocorrelation contrast and distribution distortion of wavelet coefficients. At the classification stage, a neural network approach using error back-propagation algorithm with momentum was implemented to train a multilayered perceptron (MLP) for differentiating the diagnosis of breast tumors. Using the proposed CADx system, the receiver operating characteristic (ROC) area index was 0.9396 ± 0.0183, sensitivity was 98.77%, specificity was 81.37%, with positive predictive value of 72.73% and negative predictive value of 99.24%. Considering these experimental results, the proposed diagnostic model performed very well for breast tumor diagnosis mainly because of the importance of image texture [15].

Pereira et al. [14] proposed a mutual information based method by establishing the most relevant morphometric features for discriminating breast tumors in ultrasound images. Investigation of the tumor contour in breast ultrasound images was used considering the two facts that malignant breast tumors tend to present with irregular and blurred contours, and benign tumors are usually well-defined, smooth and round. In order to rank morphometric features in discriminating breast tumors, Mutual Information and Linear Discriminant Analysis were implemented. This work revealed that among seven features investigated, the normalized residual mean square value and circularity were the most relevant descriptors for their proposed method.

Liu et al. [13] proposed a fully automated classification method for breast ultrasound images by using two steps: ROI generation and ROI classification. The first step focused on finding a credible ROI. A novel feature extraction and classification strategy was then employed by selecting some classification checkpoints that were evenly distributed in the ROI, and extraction of local texture features at every single classification checkpoint was applied. By utilizing a supervised texture classification approach in the corresponding ROI, every classification checkpoint was analyzed to classify the breast ultrasound images [13].

Yap et al. [16] carried out a comprehensive analysis of the best feature descriptors and classifiers for breast ultrasound classification. They experimented with 19 features (texture, shape and edge), 22 feature selection methods and ten classifiers. From their findings, the best combination was the feature set of 4 shape descriptors, 1 edge descriptor and 3 texture descriptors using a Radial Basis Function Network, with an area under the ROC curve of 0.948. Recently, Yap and Yap [17] conducted study to evaluate the performance of machine learning on human delineation and computer method. They found that there were no significant differences for benign lesions but computer segmentation showed better accuracy for malignant lesion classification.

Liu et al. [18] proposed a novel breast classification system for Color Doppler flow imaging and B-Mode ultrasound. In order to obtain features from B-Mode ultrasound, many feature extraction methods were used to provide both the texture and geometric features. The first stage was an extraction of color Doppler features, which was achieved by applying blood flow velocity analysis to Doppler signals to extract several spectrum features. In addition, the authors proposed a velocity coherent vector method. Furthermore, using a support

vector machine classifier, selected features were used to classify breast lesions into benign or malignant classes. They achieved an area under the ROC curve of 0.9455 when validated on 105 cases with 50 benign and 55 malignant.

There is increasing interest in deep learning for medical imaging and two research groups have been successful in using this in breast ultrasound. In 2016, Huynh et al. [19] proposed the use of a transfer learning approach for ultrasound breast images classification. The authors used 1125 cases and 2393 regions of interest for their experiment, where the ROIs were selected and labeled by the experts. To compare with the hand-crafted features, CNN was used to extract the features. When classify the CNN-extracted features with support vector machine on the recognition task of benign and malignant, they achieved an area under the ROC curve of 0.88. However, their solution was multi-stage and they did not share their dataset. In 2017, Yap et al. [20] demonstrated the use of deep learning for breast lesions detection, which outperformed the previous state-of-the-art image processing and conventional machine learning methods. They achieved an F-measure of 0.92 on breast lesions detection and made one of their datasets available for research purposes. To explore this further, we demonstrate the practicality and feasibility of using a deep learning approach for automated semantic segmentation for breast ultrasound lesion recognition.

3. METHOD

This section provides an overview of the breast ultrasound datasets, the preparation of the ground truth labeling, the proposed method and the type of performance metrics used to validate our results.

3.1 Datasets

In computer vision and medical imaging research, deep learning methods are dependent on large datasets. We combined the following two datasets to form a larger dataset for our experiment. Note that in both datasets the lesions were delineated by experienced radiologists.

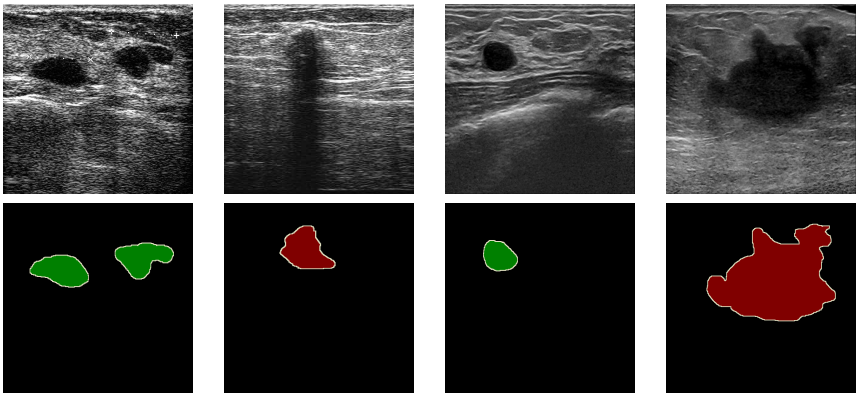
1. Dataset A: In 2001, a professional didactic media file for breast imaging specialists [21] was made available. It was obtained with B&K Medical Panther 2002 and B&K Medical Hawk 2102 US systems with an 8-12 MHz linear array transducer. Dataset A consists of 306 images from different cases with a mean image size of 377×396 pixels. From these images, 306 contained one or more lesions. Within the lesion images, 60 images presented malignant masses (as in figure 1 first row (b)) and 246 were benign lesions (as in figure 1 first row (a)). To obtain Dataset A, the user needs to purchase the didactic media file from Prapavesis et al. [21].

2. Dataset B: In 2012, the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain) has collected Dataset B with a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz). The dataset consists of 163 images from different women with a mean image size of 760×570 pixels, where the images presented one or more lesions. Within the 163 lesion images, 53 were malignant lesions (as in figure 1 first row (d)) and 110 with benign lesions (as in figure 1 first row (c)). Dataset B and the respective delineation of the breast lesions are available online for research purpose at (goo.gl/SJmoti) [20].

For a more detailed description of the datasets, please refer to [20].

3.2 Ground truth format

Since deep learning models for semantic segmentation are widely evaluated for the PASCAL-VOC 2012 training and validation dataset, these trained models are tested for various performance metrics on the PASCAL-VOC 2012 test set [22, 23]. In the PASCAL-VOC 2012 dataset, the RGB images are used as input images. The dimensions of both input images and label images should be the same size [24]. Although the images used in training are not required to be the same size for deep learning models in segmentation tasks, all the images are required to be of same size due to the use of fully connected layers in these models. In the labelled image, every pixel value for each class is an index ranging from 0 to 255. In the PASCAL-VOC 2012 dataset, there are a total of 21 classes used so far, hence, 21 indexes are used for labelling the images. For breast ultrasound images, the



(a) (b) (c) (d)

Figure 1. Original sample images (first row) and PASCAL-VOC format 8-bit paletted label images (second row).

format in digital media is generally grayscale. Hence, to make this compatible with the pre-trained models and networks that are trained for PASCAL-VOC 2012 dataset (RGB images), we converted the grayscale images to RGB images with the help of channel conversion. The ground truths in binary masks format are converted into the 8-bit paletted label images. Figure 1 illustrates the breast ultrasound images with the corresponding ground truth labeling in PASCAL-VOC format, with index 1 (RED) indicates malignant lesion and index 2 (GREEN) indicates benign lesion.

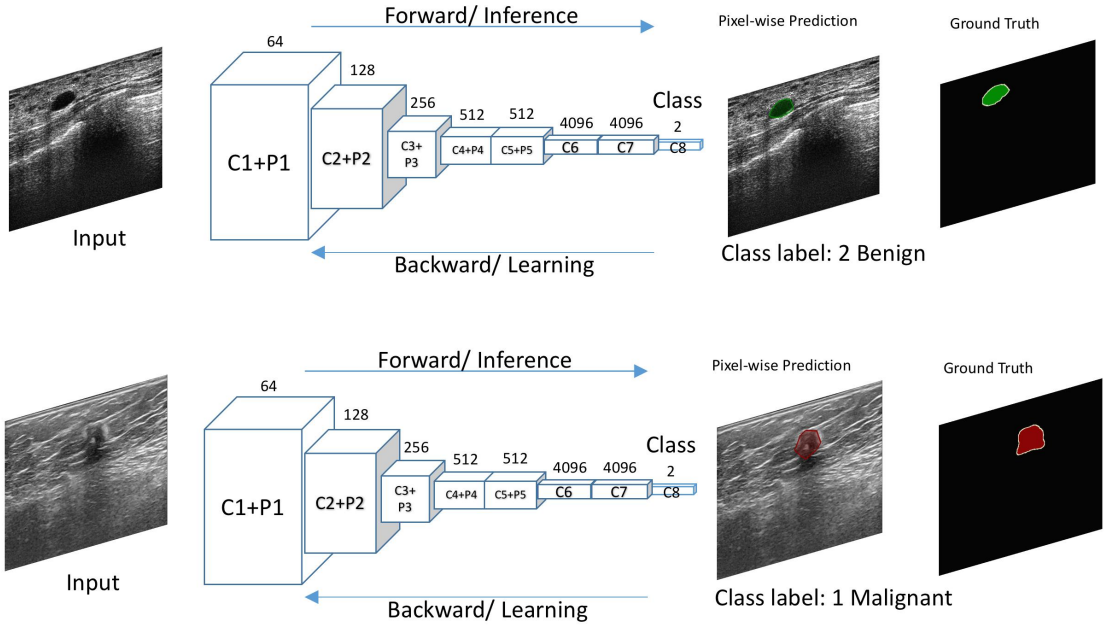


Figure 2. Illustration of end-to-end solution using semantic segmentation. The top row shows a detected benign lesion and the bottom row shows a detected malignant lesion.

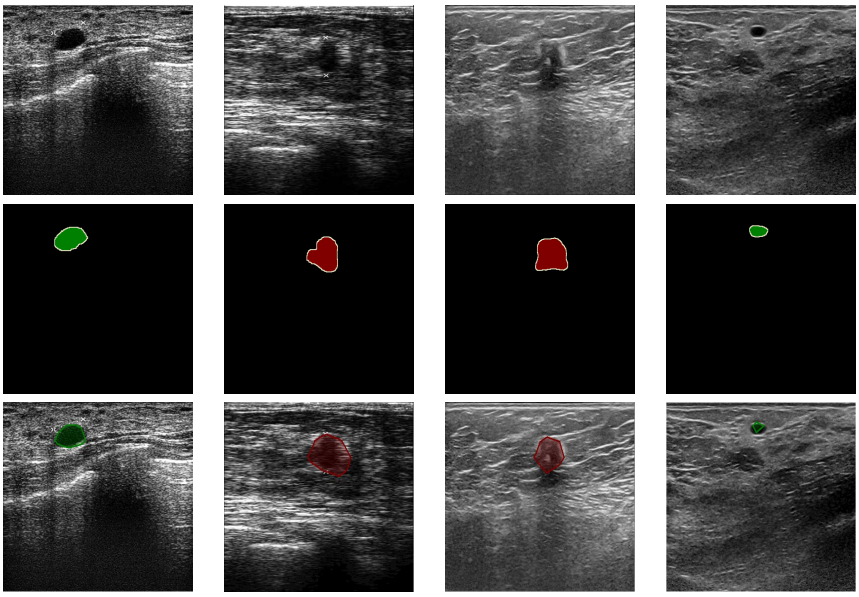
3.3 FCN-AlexNet and Transfer Learning

The deep learning methods proved its superiority over image processing methods and traditional machine learning in the detection of abnormalities in medical imaging of various modalities [20, 25]. There are two main types of tasks associated with medical imaging i.e. classification and semantic segmentation [26, 27]. But, limitation of the classification tasks is its inability to locate the abnormalities in medical imaging. Hence, semantic segmentation deep learning methods address this issues by classifying each pixel of the medical images rather than single

prediction per image in the classification task. The popular deep learning methods for end-to-end semantic segmentation are fully convolutional networks (FCNs) [28]. FCN-AlexNet is a FCN version of the original AlexNet classification model with a few adjustments in the network layers for the segmentation task [28]. This network was originally used for the classification of 1000 different objects of classes on the ImageNet dataset [29]. All deep learning frameworks rely on feature extraction through the convolution layers, but classification networks throw away the spatial information in the fully connected layers. In contrast with classification network which ignore spatial information using fully connnected layers, FCN incorporate this information by replacing fully connected layers with convolution layers. Feature maps from those convolution layers are later used for classifying each pixel to get the semantic segmentation.

Transfer Learning is a procedure where a CNN is trained to learn features for a broad domain after which layers of the CNN are fine-tuned to learn features of a more specific domain. Under this setting, the features and the network parameters are transferred from the broad domain to the specific one depending on several factors such as size of the new dataset and similarity to the original dataset. The use of deep learning methods for semantic segmentation in medical imaging suffer from the problem of data deficiency, which can be overcome with the help of transfer learning approaches [25, 26]. In this work, the pre-trained AlexNet on the ImageNet dataset which contains more than 1.5 millions images of 1000 classes was used for transfer learning [29]. The features extracted by AlexNet are transferred to the FCN-AlexNet for semantic segmentation with minor adjustments in the convolutionized fully connected layers [28].

The combination of Dataset A and Dataset B forms a larger dataset with a total of 113 malignant lesions and 356 benign lesions. We used the combined dataset to form better training and transfer learning to overcome the problem of data deficiency. We used DIGITS V5 which acts as a wrapper for the deep learning Caffe framework on the GPU machine of the following configuration: (1) Hardware: CPU - Intel i7-6700 @ 4.00Ghz, GPU - NVIDIA TITAN X 12Gb, RAM - 32Gb DDR5 (2) Deep Learning Framework: Caffe [30].



Dice: 0.9178 (a) *Dice:* 0.7194 (b) *Dice:* 0.8476 (c) *Dice:* 0.5394 (d)

Figure 3. Visual illustration of four successful cases: Original images (first row), ground truths (second row), the regions detected by our proposed method (third row) and the corresponding *Dice* score.

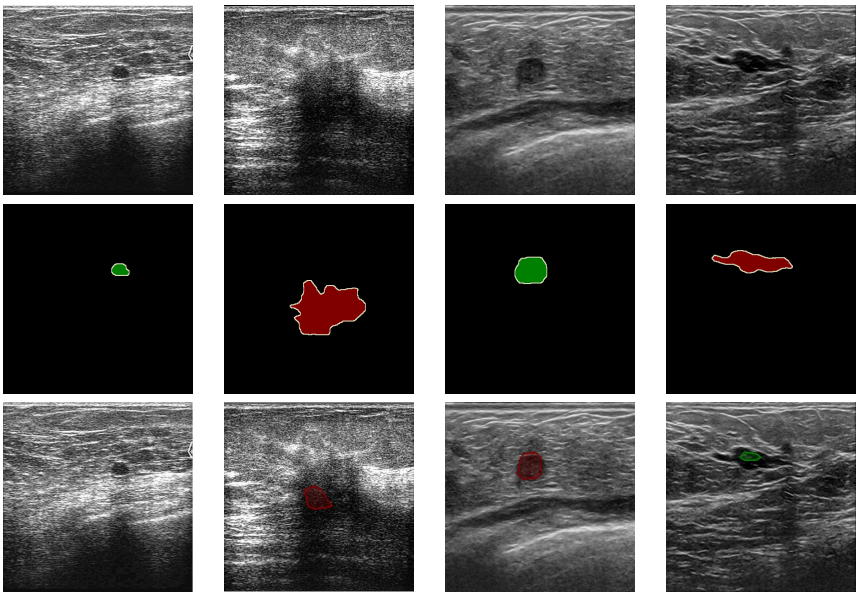
We assessed the performance of the model using the following split: 70% for training data, 10% for validation data, and 20% testing data. We trained the model using stochastic gradient descent with a learning rate of 0.0001, 60 epochs with a dropout rate of 33%. The number of epochs was kept at 60 as in [31] where convergence has already happened when we performed the empirical experiments. Figure 2 illustrates the process of the end-to-end solution using semantic segmentation.

Table 1. Summary of the performances for different lesion types for Dataset A, Dataset B and overall results using the proposed method. SD is standard deviation.

Lesion Types	Dataset	Sensitivity *Mean±SD*	Precision *Mean±SD*	Dice *Mean±SD*	MCC *Mean±SD*
Benign	A	0.7498±0.2960	0.8156±0.2544	0.7534±0.2685	0.7563±0.2596
	B	0.5533±0.3650	0.6999±0.3383	0.5455±0.3096	0.5729±0.2936
	Overall	0.6879±0.3298	0.7791±0.2862	0.6879±0.2963	0.6985±0.2821
Malignant	A	0.4476±0.3964	0.7654±0.3061	0.4588±0.3364	0.4926±0.3085
	B	0.6492±0.3183	0.7200±0.2885	0.6463±0.2841	0.6459±0.2737
	Overall	0.5484±0.3649	0.7427±0.2904	0.5525±0.3179	0.5693±0.2945

3.4 Evaluation Criteria

Even though the method is an end-to-end solution, we evaluated the results using standard performance metrics from the literature. To measure the accuracy of the segmentation results, the *Dice Similarity Coefficient (Dice)* (henceforth *Dice*) [32, 33] was used. We report our findings in *Dice*, *Sensitivity*, *Precision* and *Matthew Correlation Coefficient (MCC)* [34] as our evaluation metrics.



Dice: 0 (a) | *Dice:* 0.3046 (b) | miss-classified (c) | miss-classified (d)

Figure 4. Visual illustration of failed cases. Original images (first row), ground truth (second row), the regions detected by our proposed method (third row) and the corresponding *Dice* score.

4. RESULT AND DISCUSSION

Table 1 summarises the performance of our proposed method on Dataset A, Dataset B and the overall results. Overall, the proposed method performs better on benign lesions, with a *Dice* score of *0.6879*, compared to the malignant lesions with *Dice* score of *0.5525*.

The results show that the performance of the proposed method is dependent on the size of the dataset. In our datasets we have more benign data (356) than malignant data (113). It is also noted that the detection and recognition of benign lesions for Dataset A is better than Dataset B due to the same reasons. However, for malignant lesions detection and recognition, Dataset B has better results as the distribution of benign and malignant data being more balanced.

For breast ultrasound recognition on the combined dataset, when considering the number of images with *Dice* score > 0.5, 79.45% of the benign lesions were successfully segmented and correctly recognised, but only 65%

of the malignant lesions were successfully segmented and correctly recognised. Four examples of the successful cases for our experiment are illustrated in Figure 3 and four failed cases as illustrated in Figure 4. From our observation, the lesions with small area, ambiguity in the boundary and irregular shape are harder for semantic segmentation due to the lack of data to represent these categories.

5. CONCLUSION

The difficulties in conventional machine learning are: 1) It is based on hand-crafted features; 2) In some cases, it requires human intervention where the radiologists has to select the ROI; and 3) It is multi-stage and there is dependency from one stage to the next. In this paper, the problem was solved by using a deep learning approach where we have shown the feasibility of designing a robust end-to-end solution for breast ultrasound lesions recognition.

Although this paper has provided a new insight for future research to further investigate more on deep learning techniques, proposing an accurate end-to-end solution for breast ultrasound lesions recognition remains a challenge due to the lack of datasets to provide sufficient data representation. In the future, with the growth of big data and data sharing efforts, an end-to-end solution based on deep learning approach may find wide applications in breast ultrasound computer aided diagnosis.

REFERENCES

[1] "Breast cancer care: Facts and statistics 2017." Online. https://www.breastcancercare.org.uk/about-us/media/press-pack-breast-cancer-awareness-month/facts-statistics (Dec. 2017).

[2] Berg, W., Gutierrez, L., NessAiver, M., Carter, W. B., Bhargavan, M., Lewis, R., and Ioffe, O., "Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer," *Radiology* **233**(3), 830–849 (2004).

[3] Yap, M. H., Gale, A. G., and Scott, H. J., "Generic infrastructure for medical informatics (gimi): the development of a mammographic training system," in [*International Workshop on Digital Mammography*], 577–584, Springer (2008).

[4] Yap, M. H., Edirisinghe, E., and Bez, H., "Processed images in human perception: A case study in ultrasound breast imaging," *European Journal of Radiology* **73**(3), 682–687 (2010).

[5] Stavros, A., Rapp, C., and Parker, S., [*Breast Ultrasound*], 978-0397516247, LWW, 1 ed. (1995).

[6] Drukker, K., Giger, M. L., Vyborny, C. J., and Mendelson, E. B., "Computerized detection and classification of cancer on breast ultrasound," *Academic Radiology* **11**(5), 526–535 (2004).

[7] Yap, M. H., Edirisinghe, E., and Bez, H., "A comparative study in ultrasound breast imaging classification," *Proc.SPIE* **7259**, 7259 – 7259 – 11 (2009).

[8] Shan, J., Cheng, H., and Wang, Y., "A novel automatic seed point selection algorithm for breast ultrasound images," in [*Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*], 1–4 (2008).

[9] Cheng, H., 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).

[10] Yap, M. H., Edirisinghe, E. A., and Bez, H. E., "A novel algorithm for initial lesion detection in ultrasound breast images," *Journal of Applied Clinical Medical Physics* **9**(4), 181–199 (2008).

[11] Sehgal, C. M., Weinstein, S. P., Arger, P. H., and Conant, E. F., "A review of breast ultrasound," *Journal of Mammary Gland Biology and Neoplasia* **11**(2), 113–123 (2006).

[12] Sahiner, B., Chan, H.-P., Roubidoux, M. A., Helvie, M. A., Hadjiiski, L. M., Ramachandran, A., Paramagul, C., LeCarpentier, G. L., Nees, A., and Blane, C., "Computerized characterization of breast masses on three-dimensional ultrasound volumes," *Medical Physics* **31**(4), 744–754 (2004).

[13] Liu, B., Cheng, H., Huang, J., Tian, J., Tang, X., and Liu, J., "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images," *Pattern Recognition* **43**(1), 280 – 298 (2010).

[14] Pereira, W. C. A., Alvarenga, A. V., Infantosi, A. F. C., Macrini, L., and Pedreira, C. E., "A non-linear morphometric feature selection approach for breast tumor contour from ultrasonic images," *Computers in Biology and Medicine* **40**(11), 912–918 (2010).

[15] Chen, D., Chang, R., Kuo, W., Chen, M., and Huang, Y., "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks," *Ultrasound in Medicine and Biology* **28**(10), 1301 – 1310 (2002).

[16] Yap, M. H., Edirisinghe, E., and Bez, H., "Computer aided detection and recognition of lesions in ultrasound breast images," in [*Innovations in Data Methodologies and Computational Algorithms for Medical Applications*], 125–152, IGI Global (2012).

[17] Yap, M. H. and Yap, C. H., "Breast ultrasound lesions classification: a performance evaluation between manual delineation and computer segmentation," in [*SPIE Medical Imaging*], 978718–978718, International Society for Optics and Photonics (2016).

[18] Liu, Y., Cheng, H., Huang, J., Zhang, Y., Tang, X., Tian, J.-W., and Wang, Y., "Computer aided diagnosis system for breast cancer based on color doppler flow imaging," *Journal of Medical Systems* **36**(6), 3975–3982 (2012).

[19] Huynh, B., Drukker, K., and Giger, M., "Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks," *Medical Physics* **43**(6), 3705–3705 (2016).

[20] 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* (2017).

[21] Prapavesis, S., Fornage, B., Palko, A., Weismann, C., and Zoumpoulis, P., [*Breast Ultrasound and US-Guided Interventional Techniques: A Multimedia Teaching File*], Thessaloniki, Greece (2003).

[22] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J., "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857* (2017).

[23] Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A., "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision* **111**, 98–136 (Jan. 2015).

[24] Thoma, M., "A survey of semantic segmentation," *arXiv preprint arXiv:1602.06541* (2016).

[25] Goyal, M. and Yap, M. H., "Multi-class semantic segmentation of skin lesions via fully convolutional networks," *arXiv preprint arXiv:1711.10449* (2017).

[26] Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S., and Spragg, J., "Fully convolutional networks for diabetic foot ulcer segmentation," in [*2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*], 618–623 (Oct 2017).

[27] Goyal, M., Reeves, N. D., Davison, A. K., Rajbhandari, S., Spragg, J., and Yap, M. H., "Dfunet: Convolutional neural networks for diabetic foot ulcer classification," *arXiv preprint arXiv:1711.10448* (2017).

[28] Long, J., Shelhamer, E., and Darrell, T., "Fully convolutional networks for semantic segmentation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3431–3440 (2015).

[29] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[30] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," in [*Proceedings of the 22nd ACM international conference on Multimedia*], 675–678, ACM (2014).

[31] Duchi, J., Hazan, E., and Singer, Y., "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research* **12**(Jul), 2121–2159 (2011).

[32] Zikic, D., Ioannou, Y., Brown, M., and Criminisi, A., "Segmentation of brain tumor tissues with convolutional neural networks," *Proceedings MICCAI-BRATS* , 36–39 (2014).

[33] Pereira, S., Pinto, A., Alves, V., and Silva, C. A., "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging* **35**(5), 1240–1251 (2016).

[34] Powers, D. M., "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies* **2**, 37–63 (2011).