The 11th IEEE International Conference on E-Health and Bioengineering - EHB 2023
Grigore T. Popa University of Medicine and Pharmacy Iasi, November 9-10, 2023, Bucharest, Romania

Computer-Aided System for Breast Cancer Lesion Segmentation and Classification Using Ultrasound Images.

Saied Salem[1], Ahmed Mostafa[1], Yasien E. Ghalwash[1], Manar N. Mahmoud[1], Ahmed F. Elnokrashy[3, 4], Ahmed M. Mahmoud[1, 2]

[1]Department of Biomedical Engineering and Systems, Cairo University, Giza 12613, Egypt
saied.mahmoud00@eng-st.cu.edu.eg, ahmed.hassan002@eng-st.cu.edu.eg, yasien.ghalwsh00@eng-st.cu.edu.eg, manar.manar91@eng1.cu.edu.eg, a.ehab.mahmoud@eng1.cu.edu.eg
[2]Astute Imaging LLC, Kirkland, USA
[3]Electrical Engineering Department, Benha University, Benha, Egypt, ahmed.elnokrashy@bhit.bu.edu.eg-
[4]Computer Science Department, Faculty of Information Technology and Computer Science, Nile University, Giza 12677, Egypt, anokrashy@nu.edu.eg

***Abstract*— Breast cancer stands as a formidable global health challenge, substantially impacting cancer-related mortality rates. Ultrasound (US) imaging has gained prominence in breast cancer diagnosis, particularly for individuals with dense breast tissue. However, the efficacy of US imaging is reliant on operator proficiency and is susceptible to noise, posing a substantial diagnostic hurdle. In this study, we have devised an automated ultrasound-based Computer-Aided Diagnosis (CAD) system designed to detect and classify breast cancer lesions. A dataset comprising 6,319 images from 2889 patients was employed. To ensure the generalizability of our AI algorithm, images were acquired using various US machines with different transducers (1-14 MHz). Deep learning methodologies were harnessed, encompassing the utilization of the EfficientNetV2-B0 architecture for image classification (benign/malignant) and the implementation of the Attention U-Net coupled with the Cosh log Dice loss function for breast lesion segmentation. Our CAD system demonstrated an impressive sensitivity of 89.0% and specificity of 92.0% for classification, along with a segmentation Dice score of 86.0%. The integration of such CAD systems into breast imaging workflows holds promise for diminishing the influence of human errors, consequently reducing diagnostic costs, and expediting the breast US imaging process.**

Keywords—Breast cancer diagnosis, Ultrasound, Computer aid diagnoses (CAD), Deep Learning, Explainable AI (XAI).

I. INTRODUCTION

Breast cancer represents a significant global health challenge, being the most prevalent cancer and a leading cause of cancer-related mortality among women worldwide [1]. According to the American Cancer Society, when breast cancer is detected early, the 5-year relative survival rate is 99% [2]. Given that Full-Field Digital Mammography (FFDM) is the gold-standard in breast cancer screening since it offers the benefits of high specificity in detecting suspicious masses and microcalcifications. However, it uses radiation and is not always effective for dense breast tissue. On the other hand, Breast Ultrasound (BUS) is affordable, safe (no radiation), and good at detecting dense breast tissue [3]. Currently, the evaluation of BUS heavily relies on subjective assessments by sonographers. However, ultrasound's diagnostic precision is constrained by the availability of specialized sonographers, and even seasoned professionals often show substantial variation in their assessments. To tackle these issues, computer-aided diagnosis (CAD) systems have emerged to aid sonographers in achieving more efficient and precise breast cancer diagnoses [4][5]. The recent advancements in deep learning (DL) models have further elevated the performance of these diagnostic models, surpassing the capabilities of expert sonographers [6]. Hence, this research endeavors to create a robust CAD system for BUS imaging that can be valuable adjunct to medical professionals, increasing their diagnostic capabilities and expediting the diagnosis process to ensure timely treatment for patients.

The field of DL offers various methodologies for segmenting BUS images. Researchers have explored diverse techniques. For instance, Vakanski et al. [7] harnessed the expertise of radiologists by incorporating visual saliency maps as input to an attention block and achieved a dice score of 90.5% on a small dataset of 510 images. Shareef et al.[8] expanded upon the basic U-net by introducing dual encoders, facilitating robust segmentation for lesions of varying sizes through a row-column-wise segmentation approach and achieved a dice score of 87% on a small dataset of 562 images and 75.9% on another dataset contains 163 images. Meanwhile, Zhao et al. [9] proposed the use of ResU-net, specifically designed to mitigate the issue of vanishing gradients. They integrated an attention mechanism and employed different loss functions to optimize segmentation outcomes and achieved a dice score of 92.1%. Whereas for breast cancer classification, Mo. et al. [10] proposed a new approach to use the sequential data analysis nature of the transformer and extracts the inter- and intra-layer spatial information horizontally and vertically and achieve high results in classification of BUS in different dataset their highest result was an AUC of 92.4% on a big dataset contains 2405 images. In

979-8-3503-2887-5/23/$31.00 ©2023 IEEE

[11] Nguyen. et al investigated the efficiency of applying different augmentation techniques and used EfficientNetV2 and compared it to Fully Connected Network, shallow Convolutional Neural Networks and EfficientNetV2 model achieved an AUC of 79.8% without augmentation and 80.4% on a small dataset of 780 images.

In this study, we present a comprehensive workflow for breast lesion analysis, spanning from precise lesion segmentation to subsequent classification. Our research significantly contributes to the field by streamlining the diagnostic process with an integrated approach that ensures consistency and adherence to established standards. We further bolstered the robustness and generalizability of our models by curating a substantial and diverse breast ultrasound dataset dataset with approximately 6319 images, amalgamating data from multiple sources for rigorous training and validation. This approach equips our models to perform reliably across various clinical scenarios and patient populations, underpinning the practicality of our findings. Moreover, we introduced novel techniques, including the Log Cosh Dice loss function in tandem with the Attention U-net for lesion segmentation, which exhibited smoother convergence dynamics compared to traditional methods. Additionally, the incorporation of the EfficientNetV2-B0 model for lesion classification markedly improved model performance. Furthermore, our fine-tuning of the classification model specifically for ultrasound images represents a noteworthy advancement in breast ultrasound (BUS) image analysis, contributing to the ongoing refinement and progress of diagnostic methodologies in this critical domain.

II. Materials and Methods

A. Datasets

In this work, we used 5 different datasets with different class distribution as shown in Table I, different transducers with different acquisition frequencies, four of them are public access data and one of them is private access data. A brief description of each dataset is as follows:

TABLE I. DATASETS DISTRIBUTION

Dataset	Benign	Malignant	Normal
BUSI	210	437	133
UDIAT	109	54	-
OASBUD	104	96	-
GDPH&SYSUCC	886	1519	-
Own Dataset	2370	401	-
Total	3679	2507	133

1) BUSI [12] consists of a total of 780 BUS images from the Baheya [1] Hospital for Early Detection and Treatment of Women's Cancer, Cairo, Egypt. They were acquired with GE LOGIQ E9 using 1-5 MHz ML6-15-D Matrix linear probe. Furthermore, each image has the pixel-level ground truth of the lesion which was manually annotated by radiologists.
2) UDIAT [13] contains a total of 163 BUS images acquired with Siemens ACUSON Sequoia C512 using 8.5 MHz 17L5 HD linear array probe. All the images were collected from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell, Spain.
3) OASBUD [14] contains 200 ultrasound scans (2 orthogonal scans each) acquired by Ultrasonix SonixTouch using 5-14 MHz L14-5/38 linear array probe..
4) GDPH&SYSUCC [10] consists of a total of 2405 images, acquired by Hitachi Ascendus, Mindray DC-80, Toshiba Aplio 500 and Supersonic Aixplorer. All the images were labeled as benign or malignant according to the pathology report.
5) The last dataset were collected by our team and it consists of 2771 images acquired from RS80A, Samsung Electronics using 7-10 MHz linear probe. All the images were collected based on Institutional Review Board (IRB) with the Women and Fetal Imaging (WAFI) Center located in Cairo, Egypt [15].

B. Data Preprocessing

Ultrasound images are commonly afflicted by issues such as speckle noise [16], high variability, and complex features. Preserving essential features while eliminating noise is paramount. To address this, we employed bilateral filtering, which effectively removes noise while preserving image edges. Additionally, in some datasets, ultrasound images are encumbered by black borders. To ensure that our models are fed only pristine ultrasound images devoid of these borders, we introduced a novel algorithm that employs boundary tracing for black border removal. Subsequently, we applied adaptive histogram equalization to enhance contrast, followed by image resizing, which will be discussed further in subsequent sections. Furthermore, the datasets exhibited significant class imbalance, as depicted in Table I. To facilitate unbiased and accurate classification, we adopted four augmentation methods [17]: translation, blurring, rotation, random contrast adjustments, and gamma correction. These methods not only address class imbalance but also ensure that augmented images remain sufficiently dissimilar from the original training set, thereby mitigating the risk of model overfitting while preserving the essential properties of ultrasound images.

C. Classification

We employed a transfer learning approach, utilizing the EfficientNetV2-B0 architecture [18], which comprises a total of 270 layers. To optimize model training, we initially froze 90 layers of the pre-trained model, removed the top layer, and replaced it with two layers: a global max pooling layer and an output classification layer with softmax activation. Fine-tuning of the pre-trained model was then meticulously carried out to achieve optimal results and ensure model generalization. In our hyperparameter exploration, we conducted a systematic search and found that the most effective parameters were an input shape of 224x224, utilization of ImageNet weights, a batch size of 16, and the implementation of a decaying cyclic learning rate schedule in conjunction with the SGD optimizer. This combination yielded the best performance metrics, including AUC, specificity, cross-entropy loss, and sensitivity, as assessed on the test set.

Additionally, it was imperative to address class imbalance in the input training data to facilitate unbiased learning of correct features. Our training data consisted of datasets such as BUSI,

[1] The official Baheya website: https://baheya.org/en

UDIAT, OASBUD, GDPH, and SYSUCC. The combined dataset was split into training (80%), validation (10%), and test (10%) subsets. To balance the training dataset, we employed data augmentation techniques. The classification model was trained for a total of 14 epochs, with early stopping mechanisms in place to mitigate overfitting.

D. Segmentation

In our segmentation analysis, we employed the Attention U-net architecture with the Log Cosh dice loss function and compared the results against those obtained using the Combo loss function for lesion segmentation. Furthermore, we explored the use of U-net with various backbone networks to scrutinize the distinct characteristics of these two loss functions.

To ensure equitable comparisons, we rigorously conducted hyperparameter tuning for both loss functions. Critical hyperparameters for the segmentation model included the choice of backbone network, input shape (128x128), batch size (32), learning rate, optimizer (Adam), and the selection of the loss function (Log Cosh dice loss or Combo loss). Our objective was to identify the optimal hyperparameter configurations for each loss function to maximize model performance.

These experiments were conducted on three distinct datasets: BUSI, SYSUCC, and our proprietary dataset. To address class imbalance in the training dataset, we employed augmentation techniques. The models underwent training for a total of 75 epochs, with the weights yielding the best performance on the validation dataset being saved for further analysis.

III. Results

The performance of two loss functions including the Log Cosh dice loss and Combo loss was compared for lesion segmentation in BUS images. The U-net architecture was adopted with various backbone networks including Resnet34, DenseNet121, Efficientnetb0, and Attention U-net to assess these loss functions comprehensively.

Table II shows the Dice Similarity Scores (DSC) for different loss functions across various models. The Attention U-net with Log Cosh dice exhibited the highest DSC scores in both the validation (0.86) and test sets (0.85).

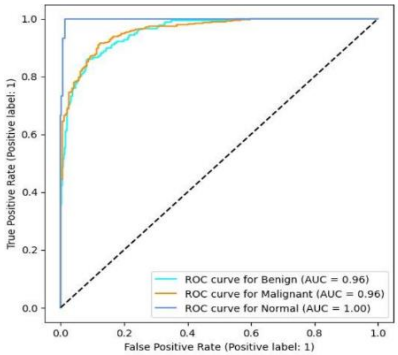

Fig. 1. ROC curve of classification model EfficientNetB0

Both models EfficientNetV2-B0 and DenseNet121 exhibited good performance to classify BUS images. As described in Table III but we decided to opt for EfficientNetV2-B0 because it is more efficient in terms of both model size and computational resources. The model exhibited a 0.87 F1 score, sensitivity of 0.89 and a specificity of 0.92, signifying its effectiveness in capturing subtle variations in BUS images. The (AUC) curve reached an exceptional value of 0.96 as shown in Fig. 1 reinforcing its robust discriminative capabilities. These results demonstrate the efficacy of the proposed classification framework. Table IV exhibits the performance of the classification model trained and tested on each dataset.

TABLE II. EVALUATION METRICS FOR SEGMENTATION

Model	*Loss*	*Val DSC*	*Test DSC*
U-net with ResNet34 backbone	Log Cosh dice	0.84	0.85
U-net with ResNet50 backbone	Combo loss	0.84	0.83
Attention U-net	Combo loss	0.85	0.84
Attention U-net	Log Cosh dice	٠.٨٦	0.85

TABLE III. EVALUATION METRICS FOR CLASSIFICATION

Model	Acc	F1 Score	Sensitivity	Specificity	AUC
VGG16	76	76	63.6	82.8	90.1
MobileNet	83	83	70.6	88	94.8
DenseNet121	89	89	89.1	92.5	97
EfficientNet B0	86	86.2	79.4	91.2	96.5
EfficientNetV 2-B0	89	87	89	92	96

TABLE IV. CLASSIFICATION RESULTS OF EFFCIENTNETV2-BO ON EACH DATASET

Dataset	Acc	F1 Score	Sensitivity	Specificity	AUC
BUSI	86.0	85.8	86.7	92.0	93.75
UDIAT	98.0	98.0	95.0	96.4	98.0
OASBUD	71.0	70.5	69.0	68.5	81.7
GDPH&SYSUCC	90.0	90.0	88.5	88.3	96.0

Furthermore, we tested the feasibility of utilizing Explainable Artificial Intelligence (XAI) to enhance the interpretability of our classification model. The Grad-CAM [19] technique was adopted to visualize the morphological features of the lesions that contributed most significantly to the model's predictions as we see in Fig. 2 the model learns to extract the irregularity of lesion geometry feature of malignant images and distinguish it from regular ellipsoid geometry lesions of benign images.

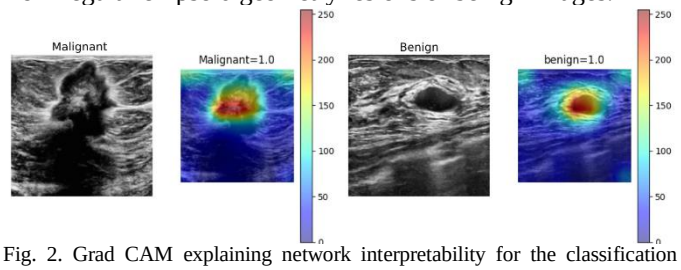

Fig. 2. Grad CAM explaining network interpretability for the classification model.

IV. Discussion

In this study, we developed a CAD system for classifying BUS images and segmenting breast lesions, which performed exceptionally well in overcoming ultrasound artifacts.

While we aimed to diversify our datasets, including variations in modalities, frequency, and operators, we

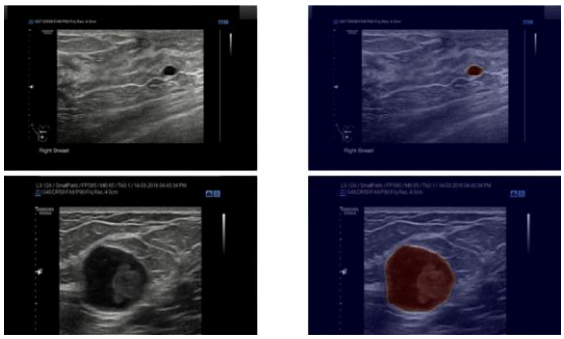

Fig. 3. Generalization of the Segmentation model output mask on different lesion shapes and size

acknowledge the need for a larger and more diverse dataset for comprehensive validation, emphasizing collaborative efforts for dataset curation. We addressed class imbalance using weighted loss functions, compared our model's performance with established ones, showing promising results in BUS image classification. XAI as shown in Fig. 2 ensured that AI-driven decisions in the medical domain are explainable and trustworthy. As we see our model exhibited robustness in distinguishing lesion boundaries from acoustic shadowing. We evaluated our model on both combined and individual datasets. Notably, it performed well on all datasets, with the highest accuracy observed on high-resolution datasets (Table IV). These results confirm the model's capacity for unbiased generalization across diverse datasets.

We also compared two loss functions for breast lesion segmentation, utilizing U-net architecture with various extensions. Both Log Cosh dice loss and Combo loss performed well, with a slight advantage observed in the Attention U-net with Log Cosh dice loss showing that the model effectively generalizing over the diverse characteristics of the lesion as shown in Fig. 3.

We recognized limitations related to training data variations and recommended standardized annotation protocols for more accurate ground truth. Overall, our work highlighted the importance of data sharing, quality assurance, and model adaptability in advancing breast ultrasound image analysis. Future research should focus on addressing these challenges to improve clinical outcomes in breast cancer assessment using BUS.

V. Conclusions

In conclusion, we presented a comprehensive exploration of breast ultrasound image analysis, encompassing both classification and lesion segmentation. For classification, our EfficientNetV2-B0-based model has demonstrated exceptional accuracy and diagnostic performance, outperforming established models. Holding promises for improving the early detection of breast cancer. In the context of lesion segmentation, our approach has shed light on the importance of selecting an appropriate loss function. Log Cosh dice loss, in conjunction with the Attention U-net model, emerged as a powerful combination for capturing lesion areas accurately. These contributions advance the field of breast ultrasound image analysis, and we anticipate that they will pave the way for more efficient breast cancer diagnosis methods.

VI. Acknowledgment

We would like to express our gratitude to Astute Imaging for providing the technical and research support throughout this work.

VII. References

[1] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021.

[2] A. N. Giaquinto *et al.*, “Breast Cancer Statistics, 2022,” *CA Cancer J Clin*, vol. 72, no. 6, pp. 524–541, Nov. 2022.

[3] T. B. Bevers *et al.*, “Breast cancer screening and diagnosis, version 3.2018, NCCN clinical practice guidelines in oncology,” *Journal of the National Comprehensive Cancer Network*, vol. 16, no. 11, pp. 1362–1389, 2018.

[4] M. H. Yap, E. Edirisinghe, and H. Bez, “Processed images in human perception: A case study in ultrasound breast imaging,” *Eur J Radiol*, vol. 73, no. 3, pp. 682–687, Mar. 2010.

[5] Z. Guo *et al.*, “A review of the current state of the computer-aided diagnosis (CAD) systems for breast cancer diagnosis,” *Open Life Sci*, vol. 17, no. 1, pp. 1600–1611, Jan. 2022.

[6] X. Qian *et al.*, “A combined ultrasonic B-mode and color Doppler system for the classification of breast masses using neural network,” *Eur Radiol*, vol. 30, no. 5, pp. 3023–3033, May 2020.

[7] A. Vakanski, M. Xian, and P. Freer, “Attention Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images,” *Ultrasound Med Biol*, vol. 46, no. 10, pp. 2819–2833, Oct. 2019.

[8] B. Shareef, A. Vakanski, P. E. Freer, and M. Xian, “ESTAN: Enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation,” *Healthcare 2022, Vol. 10, Page 2262*, vol. 10, no. 11, p. 2262, Nov. 2022.

[9] T. Zhao and H. Dai, “Breast Tumor Ultrasound Image Segmentation Method Based on Improved Residual U-Net Network,” *Comput Intell Neurosci*, vol. 2022, 2022.

[10] Y. Mo *et al.*, “HoVer-Trans: Anatomy-aware HoVer-Transformer for ROI-free Breast Cancer Diagnosis in Ultrasound Images,” *IEEE Trans Med Imaging*, vol. 42, no. 6, pp. 1696–1706, May 2022.

[11] H. T. Nguyen, L. N. Le, T. M. Vo, D. N. T. Pham, and D. T. Tran, “Breast Ultrasound Image Classification Using EfficientNetV2 and Shallow Neural Network Architectures,” *Lecture Notes in Networks and Systems*, vol. 497 LNNS, pp. 130–142, 2022.

[12] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data Brief*, vol. 28, p. 104863, Feb. 2020.

[13] Y. MH *et al.*, “Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks,” *IEEE J Biomed Health Inform*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.

[14] H. Piotrzkowska-Wróblewska, K. Dobruch-Sobczak, M. Byra, and A. Nowicki, “Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions,” *Med Phys*, vol. 44, no. 11, pp. 6105–6109, Nov. 2017.

[15] M. Mahmoud, M. Rushdi, I. Ewais, E. Hosny, H. Gewefel, and A. Mahmoud, “Computationally-efficient wavelet-based characterization of breast tumors using conventional B-mode ultrasound images,” p. 88, Mar. 2019.

[16] M. Moinuddin, S. Khan, A. U. Alsaggaf, M. J. Abdulaal, U. M. Al-Saggaf, and J. C. Ye, “Medical ultrasound image speckle reduction and resolution enhancement using texture compensated multi-resolution convolution neural network,” *Front Physiol*, vol. 13, p. 961571, Nov. 2022.

[17] B. M. Shareef *et al.*, “A Benchmark for Breast Ultrasound Image Classification”.

[18] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” *Proc Mach Learn Res*, vol. 139, pp. 10096–10106, Apr. 2021, Accessed: Sep. 12, 2023.

[19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Oct. 2016.