Ultrasonics 121 (2022) 106682

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Joint segmentation and classification of breast masses based on ultrasound radio-frequency data and convolutional neural networks

Michal Byra [a,*], Piotr Jarosik [a], Katarzyna Dobruch-Sobczak [a,b], Ziemowit Klimonda [a], Hanna Piotrzkowska-Wroblewska [a], Jerzy Litniewski [a], Andrzej Nowicki [a]

[a] Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
[b] Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland

ARTICLE INFO

Keywords:
Breast mass classification
Breast mass segmentation
Convolutional neural networks
Deep learning
Quantitative ultrasound
Ultrasound imaging

ABSTRACT

In this paper, we propose a novel deep learning method for joint classification and segmentation of breast masses based on radio-frequency (RF) ultrasound (US) data. In comparison to commonly used classification and segmentation techniques, utilizing B-mode US images, we train the network with RF data (data before envelope detection and dynamic compression), which are considered to include more information on tissue's physical properties than standard B-mode US images. Our multi-task network, based on the Y-Net architecture, can effectively process large matrices of RF data by mixing 1D and 2D convolutional filters. We use data collected from 273 breast masses to compare the performance of networks trained with RF data and US images. The multi-task model developed based on the RF data achieved good classification performance, with area under the receiver operating characteristic curve (AUC) of 0.90. The network based on the US images achieved AUC of 0.87. In the case of the segmentation, we obtained mean Dice scores of 0.64 and 0.60 for the approaches utilizing US images and RF data, respectively. Moreover, the interpretability of the networks was studied using class activation mapping technique and by filter weights visualizations.

1. Introduction

Breast cancer is the most frequent cancer among women worldwide [1]. Ultrasound (US) imaging is commonly used by diagnosticians for breast mass characterization. US imaging is inexpensive and accessible, however, analysis of B-mode US images is time-consuming and associated with high inter-rater reliability due to large variations in breast mass US image characteristics. Various computer aided diagnosis systems have been proposed to aid the radiologists in breast mass diagnosis, which is commonly related to the problems of breast mass segmentation and classification [2]. The aim of the segmentation models is to help accurately outline breast mass regions, while the classification methods usually aid with the differentiation of malignant and benign breast masses.

Methods for breast mass diagnosis are usually developed using B-mode US images. In comparison to US images, radio-frequency (RF) US data (data before envelope detection and dynamic compression) include information related to tissue physical properties coded in amplitude and phase of backscattered RF signals [3]. While B-mode US images can be used to localize tissues and assess relative echogenicity, the information on specific tissue physical properties is only partially present in US images due to lossy compression necessary to reconstruct US images and make them visible to human eye. Quantitative US (QUS) methods have been proposed to utilize RF data to extract parameters related to various physical properties of tissues, such as the backscattering or attenuation coefficients [4]. Stochastic modeling of RF data has been applied to determine local spatial distribution of tissue microstructures [5]. These methods, for example based on the Nakagami and Homodyned K distributions, have been successfully applied for breast mass classification in several papers [6–11].

Nowadays, deep learning methods based on convolutional neural networks (CNNs) are gaining momentum in breast mass segmentation and classification (see the Related Works section). In comparison to standard approaches to image recognition, requiring feature engineering, deep learning algorithms can automatically process input data to extract efficient features for recognition. However, the usefulness of machine learning methods for the processing of RF data has not yet been fully understood. QUS techniques, originating from US physics, have been devised based on specific tissue models [3]. In practice, researchers have to estimate many different QUS parameters and determine the useful ones for the investigated tissue characterization problem. Data driven machine learning models have the possibility to

* Corresponding author.
E-mail address: mbyra@ippt.pan.pl (M. Byra).

https://doi.org/10.1016/j.ultras.2021.106682
Received 1 February 2021; Received in revised form 8 December 2021; Accepted 30 December 2021
Available online 14 January 2022

directly extract information from RF signals and automatically provide useful features for tissue characterization. Uniyal et al. proposed a machine learning approach to analysis of small 2D patches of RF data manually extracted from breast mass area [12]. The authors extracted various handcrafted features, like the Higuchi fractal dimension, and used them to classify breast masses. Jarosik et al. proposed to utilize CNNs to automatically process 2D patches of RF data manually extracted from breast masses to differentiate malignant and benign masses [13]. Although not directly related to breast mass characterization, Han et al. used US RF data to develop a deep learning model for liver fat level assessment in patients with non-alcoholic fatty liver disease [14]. The authors utilized 1D RF signals manually extracted from liver area to develop 1D CNNs for the investigated tasks. Sanabria et al. used small patches of 2D RF data manually extracted from human liver tissue to assess liver fat [15]. Nguyen et al. developed 1D CNNs based on RF data collected from rabbits to quantify liver fat [16]. Interestingly, the authors presented that the RF based network outperformed QUS techniques. Moreover, RF data have been utilized for displacement estimation in quasi-static and dynamic US elastography [17–20].

The main aim of this work is to utilize RF data to develop a deep learning model for breast mass diagnosis. In comparison to the previous approaches, utilizing small patches of RF data, we propose a method that can automatically process large 2D volumes of RF data. Moreover, the proposed model can jointly perform breast mass segmentation and classification. In literature, these two tasks are commonly investigated separately, but they are in fact closely connected. Gomez-Flores et al. presented in a review paper that the shape descriptors are the better performing features for breast mass classification [21]. This suggests that a deep learning model developed for the automatic breast mass segmentation should be also able to provide shape features for differentiation of malignant and benign masses. In this work, we use Y-Nets, a modification of the popular U-Net segmentation CNN, to perform joint segmentation and classification of breast masses [22,23]. We develop the multi-task model based on RF data, and compare it with the model trained using US images. Moreover, previous papers utilizing RF data for deep learning based tissue characterization did not investigate the problem of model interpretability. Here, we also present several insights suggesting how the RF based deep learning models conduct decisions.

2. Related works

Methods for the joint segmentation and classification have been utilized in literature, for example for microscopy images and liver US images [23,24]. Nevertheless, to the best of our knowledge, this is the first paper addressing the problem of joint segmentation and classification of breast masses with deep learning methods based on US RF data. Below we describe the previous contributions related to the problems of mass segmentation and classification based on US images.

2.1. Classification networks

Several deep learning methods have been proposed for breast mass classification, commonly based on transfer learning with networks pre-trained on the ImageNet dataset [25]. In this case, pre-trained networks such as the VGG19 or InceptionV2 were used to either perform fine-tuning or extract features for classification [26–31]. Qi et al. proposed a region enhance mechanism to help better localize masses in US images and improve classification accuracy [32]. Moon et al. developed an ensemble of neural networks for mass classification [33]. Zhang et al. proposed a deep learning based approach to mass classification utilization both US images and Breast Imaging-Reporting and Data System (BI-RADS) categories [34]. Cao et al. investigated negative effects of noisy labels on deep learning based mass classifiers [35]. Similarly, Byra et al. investigated the robustness of deep learning based mass classifiers to adversarial attacks [36].

2.2. Segmentation networks

Various deep learning models, usually based on fully convolutional networks (FCN), have been devised for breast mass segmentation. Yap et al. proposed a FCN model for mass segmentation in US images [37]. To improve the performance, the authors utilized weights from VGG network trained on the ImageNet dataset. Moreover, Yap et al. also investigated the usefulness of different deep learning models for breast mass detection [38,39] Xu et al. compared the performance of FCN, U-Net, and dilated residual network in breast mass classification [40]. Similarly, Gomez-Flores et al. investigated the usefulness of several deep learning models, including U-Net and FCNs, for breast mass segmentation. Byra et al. proposed a selective kernel U-Net CNN, a modified version of the U-Net, to take into account large variability in mass sizes in automatic segmentation [41]. Han et al. proposed a semi-supervised breast mass segmentation model [42]. The author used generative adversarial networks to improve performance of FCNs and improve segmentation performance.

3. Materials and methods

3.1. Ultrasound data

This retrospective study was approved by the Institutional Review Board. RF data were collected by an experienced radiologist using Ultrasonix research US scanner (Ultrasonix Medical Corporation, Canada) equipped with L14-5/38 linear probe. Standard beamforming was applied to acquire US data, with the focal point set on breast mass area, We collected 546 RF data matrices from 273 breast masses (one mass per scan, two perpendicular scans per mass). 124 masses were malignant and 149 masses were benign. All masses were assessed either by biopsy or a two year follow up in the case of the benign masses. Dimensions of each RF data matrix were equal to 2048×256 (42 mm $\times$ 38 mm), corresponding to 256 RF signal scan lines sampled at 40 MHz. Imaging pulse center frequency was equal to around 6 MHz.

Breast mass B-mode US images were reconstructed based on RF data. First, RF data amplitude was computed using Hilbert transform. Second, amplitude samples were logarithmically compressed and mapped to gray scale US image pixel intensities (8 bits) using typical threshold level of 50 dB. To generate US images, the compressed and thresholded data, size of 2048×256, were resized to 256×256 using bi-cubic interpolation method, and processed with a 3×3 median filter. Fig. 1 presents several US images of malignant and benign breast masses. US image reconstruction scheme is depicted in Fig. 2. In the next step, the reconstructed US images were used by the radiologist to manually outline regions of interest (ROIs) presenting breast masses.

3.2. Deep learning methods

Architecture of the proposed Y-Net is illustrated in Fig. 3. U-Net segmentation CNN, consisting of contraction (encoder) and expansion (decoder) paths, served as the backbone for the multi-task network. In a standard U-Net, input image is first compressed with convolutional layers to a compact representation (central block), and used to generate segmentation mask with transposed and regular convolutional layers. Additionally, skip connections are used to concatenate feature maps from the contraction and expansion blocks to improve the training. In the case of the Y-Net architecture, U-Net is equipped with an additional classification branch, utilizing features from the central block of the U-Net [23]. In our case, we also decided to include features extracted from the expansion path. The rationale for this was that these features should include information related to breast mass shape, presumably useful for the classification. Moreover, this way we also utilized features from the contraction path propagated through the skip connections. To extract features for classification, global average pooling was applied. Next, the features were concatenated and a fully connected layer with a sigmoid

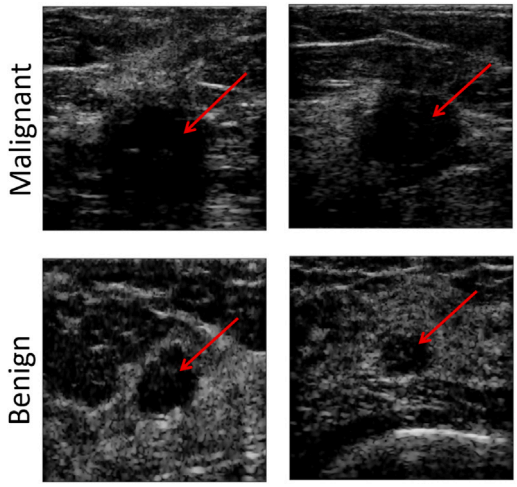

Fig. 1. US images presenting malignant and benign breast masses (red arrows).

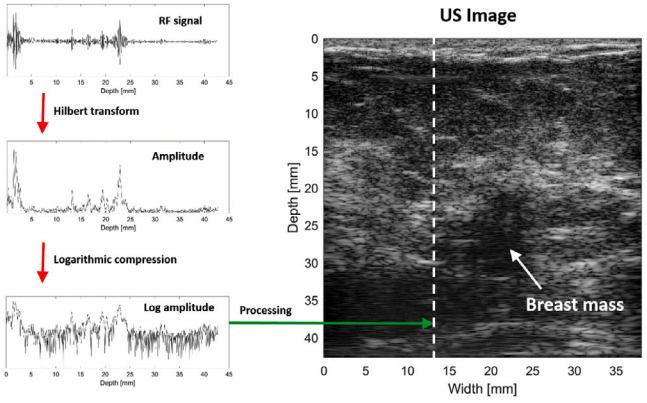

Fig. 2. General scheme presenting reconstruction of an ultrasound image scan line based on radio-frequency signal.

activation function was used to calculate *a posteriori* probability related breast mass malignancy. Each convolutional block (CB) of the Y-Net included two sub-blocks, each consisting of a 3×3 convolutional filters, batch-normalization layer and rectified linear unit (ReLU) activation function. The first CB included 16 convolutional filters, the number of filters was doubled with every block to the maximal number of 256 filters in the central block. For the blocks of the expansion path, the number of filters were consequently divided by two with each CB.

In this work, we developed two Y-Nets. Input and output dimensions of the first network were equal to 256×256, and this model was trained using US images, manual segmentations and malignant/benign labels. The second network was trained based on RF data. However, to be able to process 2048×256 RF data matrices, we additionally equipped the network with a STEM layer, which aim was to down-sample RF data in longitudinal direction with 1D convolutional filters, see Fig. 3. The first block of the STEM included 32 convolutional filters (stride parameter equal to 2, size of 48×1 corresponding to length of around 1 mm) followed by a batch-normalization layer, ReLU function and 2×1 max-pooling layer. The second block included 32 convolutional filters (stride parameter set to 1, size of 24×1) followed by a batch-normalization layer, ReLU function and 2×1 max-pooling layer. The output of the STEM block, size of $256 \times 256 \times 32$ was next processed by with the blocks corresponding to the Y-Net contraction path.

3.3. Training and evaluation

The dataset of 273 breast masses was randomly divided into training, validation and test sets with a 150/41/82 split. Horizontal flipping was applied to generate more data for training. The ratio of malignant and benign masses was maintained for each set. The Y-Net was trained to jointly minimize segmentation and classification loss, given as follows:

$$J(A, M, p, c) = J_{Dice}(A, M) + \alpha J_{clas}(p, c), \tag{1}$$

where A, M, p, c are the automatic segmentation, manual segmentation, probability score (Y-Net's classification branch) and reference label (malignant/benign). J_{clas} stands for the standard binary cross-entropy loss, α is the weighting factor and J_{Dice} is the Dice score based loss function defined in the following way:

$$J_{Dice}(A, M) = 1 - \text{Dice}(A, M), \tag{2}$$

where $\text{Dice}(A, M)$ indicates the soft Dice score [43,44]. Both networks were trained using Adam optimization method to minimize the loss given by Eq. (1) [45]. Optimal learning rates were selected based on the validation set. The weighting factor α was set to 0.5, as determined based on the validation set. To address the class imbalance, we additionally weighted the binary cross-entropy loss with weights inversely proportional to class frequencies in the training set. Batch size was equal to 12. The learning rate was exponentially decreased every 4 epochs by a factor of 0.9 if no improvement was observed on the validation set, and the training was stopped if no improvement in respect to the loss was observed on the validation set after 15 epochs. Training was repeated five times for each network, and the better-performing model on the validation set was selected for further evaluation. Segmentations calculated for the test set were additionally processed. First, morphological operators were applied to fill the holes in binary ROIs. Second, the largest ROI was selected as the candidate for the mass region.

We used standard metrics to evaluate performance on the test set. In the case of the classification task, we calculated the receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC). Accuracy, sensitivity and specificity values were calculated based on the point on the ROC curve closest to curve upper left corner [46]. DeLong test was applied to compare AUCs obtained for the models trained with RF data and US images [47]. In the case of the segmentation task, we calculated the Dice scores, pixel-level accuracy and detection rate as the ratio of correctly detected breast masses. Breast mass was considered correctly detected if the centroid of the automatic ROI was within the manual ROI. Additionally, we used the metric proposed by Yap et al. and calculated mean Dice score for the cases that obtained Dice score above 0.5 (Dice > 0.5) [38]. Standard deviations of all employed metrics were calculated with the bootstrap technique. Calculations were performed in Matlab (Mathworks, USA) and in Python using TensorFlow [48]. The networks were trained on a computer equipped with a GeForce RTX 2080 Ti graphics card.

3.4. Interpretability

To better understand how the proposed network conducts decisions, we performed two experiments. First, following the training, we visualized the weights of the first convolutional filters of the STEM, which were directly used to process RF data. We applied the Fourier transform to calculate the mean frequency of each convolutional filter. Second, we generated class activation maps (CAMs) based on feature maps utilized by the Y-Net classification branch [49,50]. We qualitatively analyzed the maps, examined what regions were highlighted by the technique, and also compared CAMs generated for the RF data and US image based Y-Nets. Feature maps extracted from each block of the Y-Net (see Fig. 3) were resized to match US image dimension, and weighted using classification layer weights to yield the activation maps.

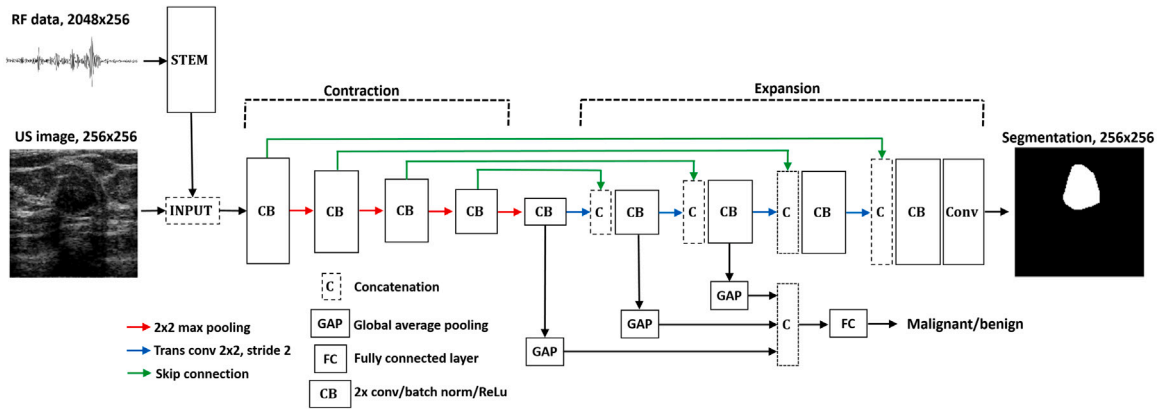

Fig. 3. Architecture of the Y-Net convolutional neural network used for joint classification and segmentation of breast masses. In our study, we compared performance of Y-Nets trained using two different ultrasound data types: ultrasound images and radio-frequency data. The only difference between the networks was that the STEM layer was used in the case of the RF data to perform down-sampling in the longitudinal direction.

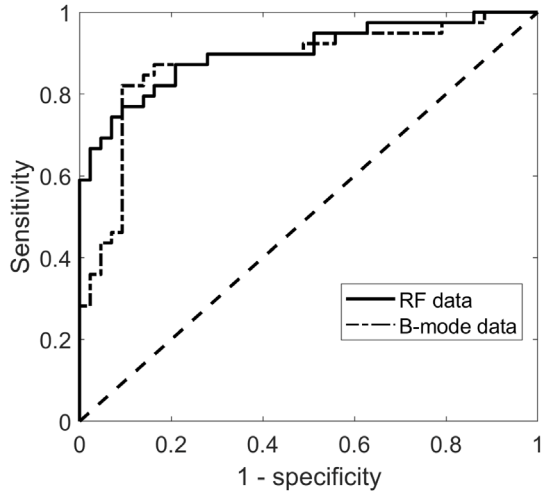

Fig. 4. Receiver operating characteristic (ROC) curves presenting good performance of the developed breast mass classification methods. The areas under the ROC curves were equal to 0.874 and 0.901 for the networks trained using US images and RF data, respectively.

4. Results

4.1. Classification

Classification performance is presented in Table 1. Generally, we obtained good performance for both deep learning methods. Network trained with RF data achieved higher AUC value, 0.901, than the model utilizing reconstructed US images, AUC of 0.874. However, while the AUC value was higher for RF data based network, DeLong test showed that there was no statistical difference between the models (p-value >0.05). ROC curves calculated for both methods are depicted in Fig. 4.

4.2. Segmentation

Table 2 presents segmentation scores obtained for each network. The model trained with US images generally achieved better results, Dice score of 0.644, than the RF data based model, Dice score of 0.601. Similarly, the US image based model achieved higher detection rate of breast masses. However, when it comes to the correctly detected breast masses, the performance of both models was similar, which is illustrated by approximately the same values of the median Dice scores and Dice > 0.5 metric. Moreover, although the detection rates were higher for the malignant masses, we found that the automatic segmentation of the malignant masses was generally more difficult for the networks, which is depicted by lower median Dice scores in Table 2.

Representative segmentations obtained for each network are presented in Fig. 5. We found that automatic segmentations were usually more smooth than the reference ROIs prepared by the radiologist. In contrary, Fig. 6 presents cases for which the networks failed to provide good segmentations, for example due to the presence of shadowing artifacts.

4.3. Interpretability

Weights of several 1D filters from the first convolutional layer of the STEM block are depicted in Fig. 7. To process RF signals, the network developed filters presenting different oscillating patterns. Additional analysis, Fig. 8, showed that the average frequencies of these filters were close to the center frequency of the imaging pulse, equal to 6 MHz. This result suggests that the network learned to decompose RF signals in the frequency domain to process RF data. Fig. 9 presents CAMs obtained for test cases classified with high confidence, *a posteriori* probability of malignancy above 0.8 and below 0.2 for the malignant and benign masses, respectively. We found that for these examples the networks worked in a similar way, both highlighting approximately the same image areas. For the malignant masses, the region of strong positive activation overlapped with the breast mass area and its surroundings. For the benign cases, the networks similarly highlighted mass area in CAMs, but in these cases the positive activation was lower and accompanied by higher negative activation.

5. Discussion

We presented that US RF data can be used for the joint classification and segmentation of breast masses. In comparison to the previous papers utilizing RF data and machine learning methods for mass classification, our approach could automatically process large volumes of RF data to yield decisions. We equipped the Y-Net with a STEM layer that effectively down-sampled RF data with 1D convolutional filters. In our study, the RF data based network provided similar AUC score, 0.901, to the B-mode US image based model, 0.874. Due to the differences in approaches and datasets, it is difficult to directly compare this result to those from the previous papers on machine learning models utilizing RF data. Method proposed by Uniyal et al. utilized small 2D patches of RF data extracted from mass area to calculate features and train support vector machine classifiers [12]. The study was performed based on a set of temporal sequences of RF data collected from 22 subjects with the same research US scanner as in the case of our work. The better performing approach presented by the authors achieved AUC value of

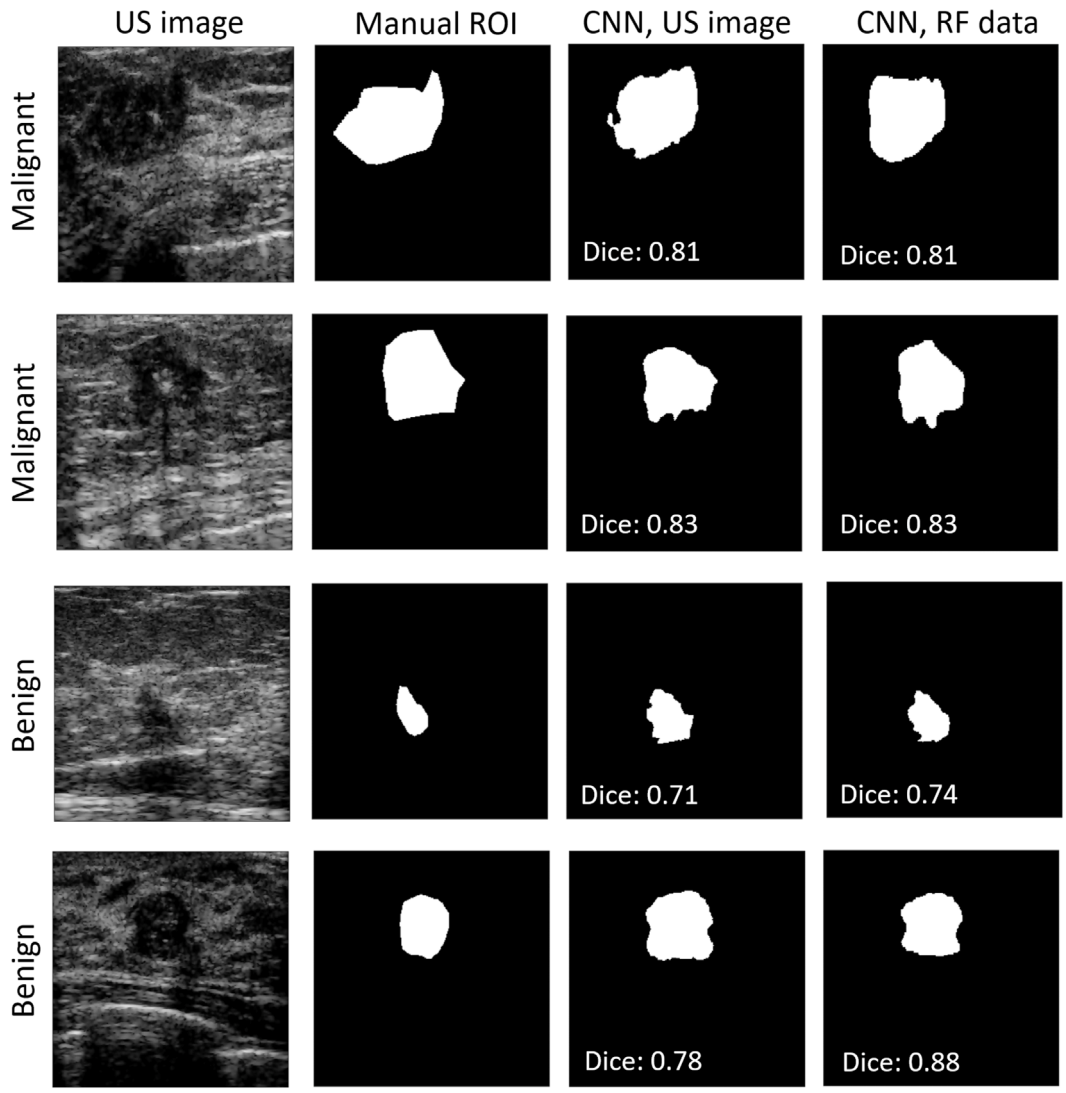

Fig. 5. Representative segmentation results (Dice score around test set median) obtained with the networks trained using US images and RF data.

Table 1
Breast mass classification performance scores (plus standard deviation) achieved by the Y-Net on test US images and RF data.

Training data type	AUC	Accuracy	Sensitivity	Specificity
US images	0.874 (±0.034)	0.865 (±0.033)	0.820 (±0.043)	0.906 (±0.042)
RF data	0.901 (±0.027)	0.829 (±0.029)	0.820 (±0.047)	0.837 (±0.045)

Table 2
Breast mass segmentation performance scores (plus median and standard deviation) achieved by the Y-Net on test US images and RF data.

Training data type	Mass type	Dice	Dice > 0.5	Accuracy	Detection rate
US images	Benign	0.660 (0.786 ± 0.300)	0.795 (0.818 ± 0.114)	0.950 (0.962 ± 0.040)	0.791
	Malignant	0.626 (0.705 ± 0.263)	0.746 (0.758 ± 0.115)	0.901 (0.914 ± 0.074)	0.846
	All	0.644 (0.747 ± 0.283)	0.772 (0.795 ± 0.117)	0.927 (0.947 ± 0.064)	0.817
RF data	Benign	0.602 (0.805 ± 0.354)	0.819 (0.838 ± 0.102)	0.964 (0.979 ± 0.048)	0.721
	Malignant	0.599 (0.679 ± 0.277)	0.748 (0.755 ± 0.109)	0.920 (0.939 ± 0.068)	0.846
	All	0.601 (0.746 ± 0.319)	0.785 (0.812 ± 0.111)	0.943 (0.967 ± 0.062)	0.780

0.82. In the previous paper from our group, Jarosik et al. built on the method proposed by Uniyal et al. and used convolutional networks to process small 2D patches of RF data and classify breast masses [13]. The networks were developed using the OASBUD, a publicly available subset of our dataset, including RF data from 100 subjects [51]. The better performing model achieved AUC value of 0.772. Therefore, the model presented in this paper achieved better performance than the previous methods, but was also developed using a larger set of RF data.

In our study we compared the performance of RF data and US images based models. However, detailed comparisons between different classification and segmentation methods based on US images was beyond the scope of this work, especially given the small volume of our dataset. Nevertheless, we can compare the results presented in previous papers on US images based models with ours. Several deep learning based approaches have been proposed for breast mass classification in US images. Authors utilized different transfer learning

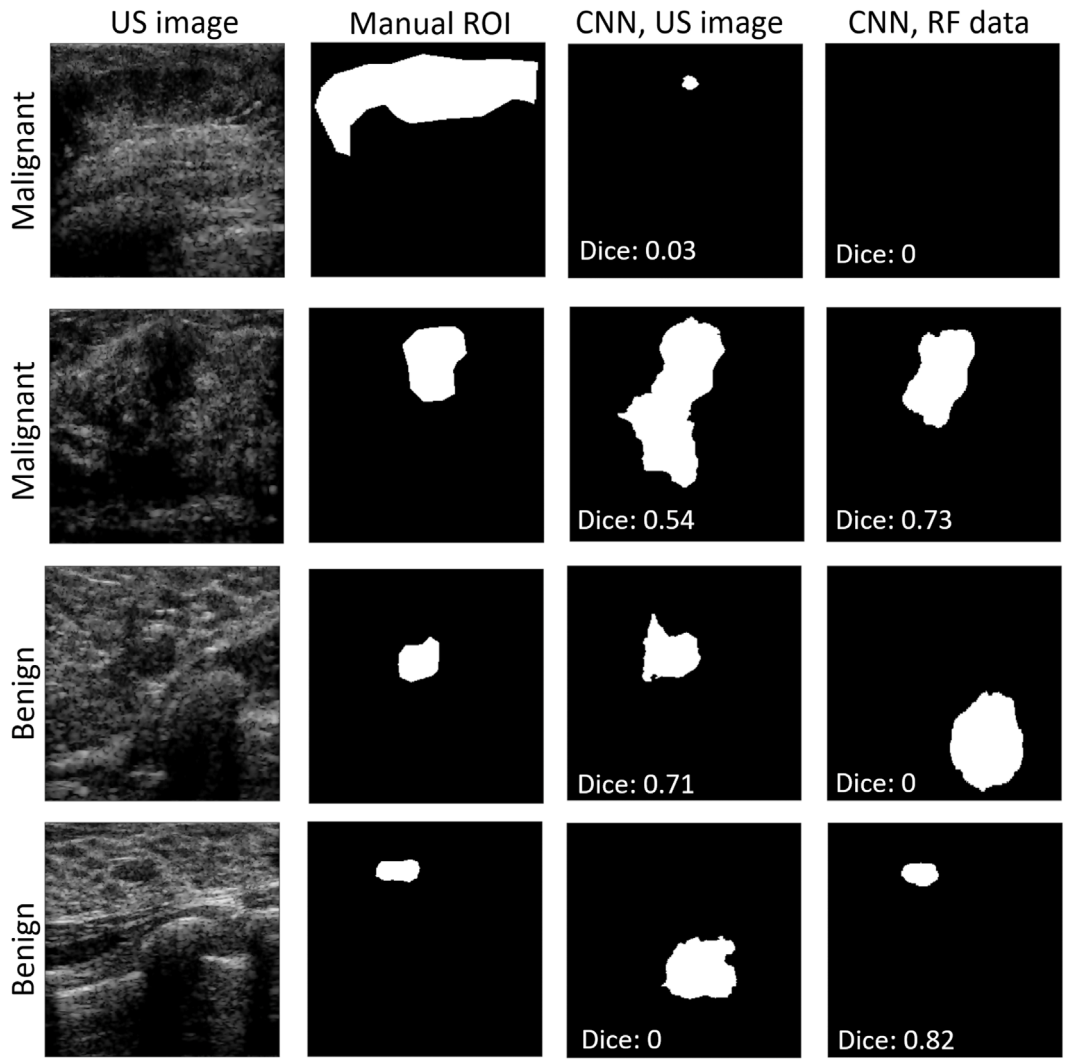

Fig. 6. Test set examples presenting poor segmentation performance. In the case of the first malignant mass, the networks presumably failed to segment the mass due to its large area and shallow position. In the case of the second mass, the network trained using ultrasound images over segmented the mass due to the shadowing artifact. Ultrasound images of the benign masses present cases where the networks missed the masses and segmented deeper dark areas instead.

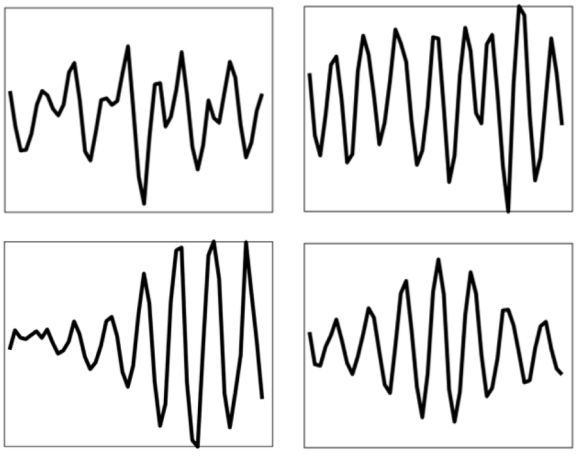

Fig. 7. Weights of the 1D convolutional filters (48 × 1) from the first layer of the network trained to process RF data. Filters present different oscillating patterns.

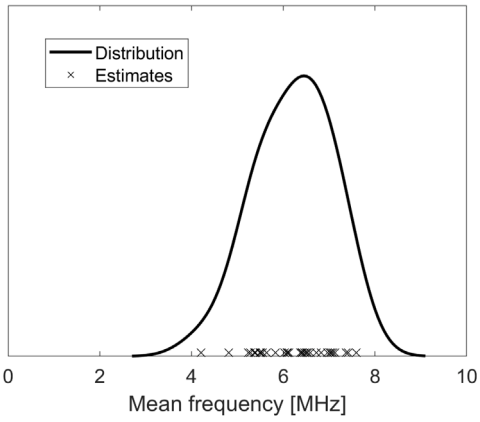

Fig. 8. Distribution of mean frequencies of the 1D convolutional filters (48 × 1) from the first layer of the network trained to process RF data. Mode of the distribution approximately matched the center frequency of the imaging pulse, equal to around 6 MHz.

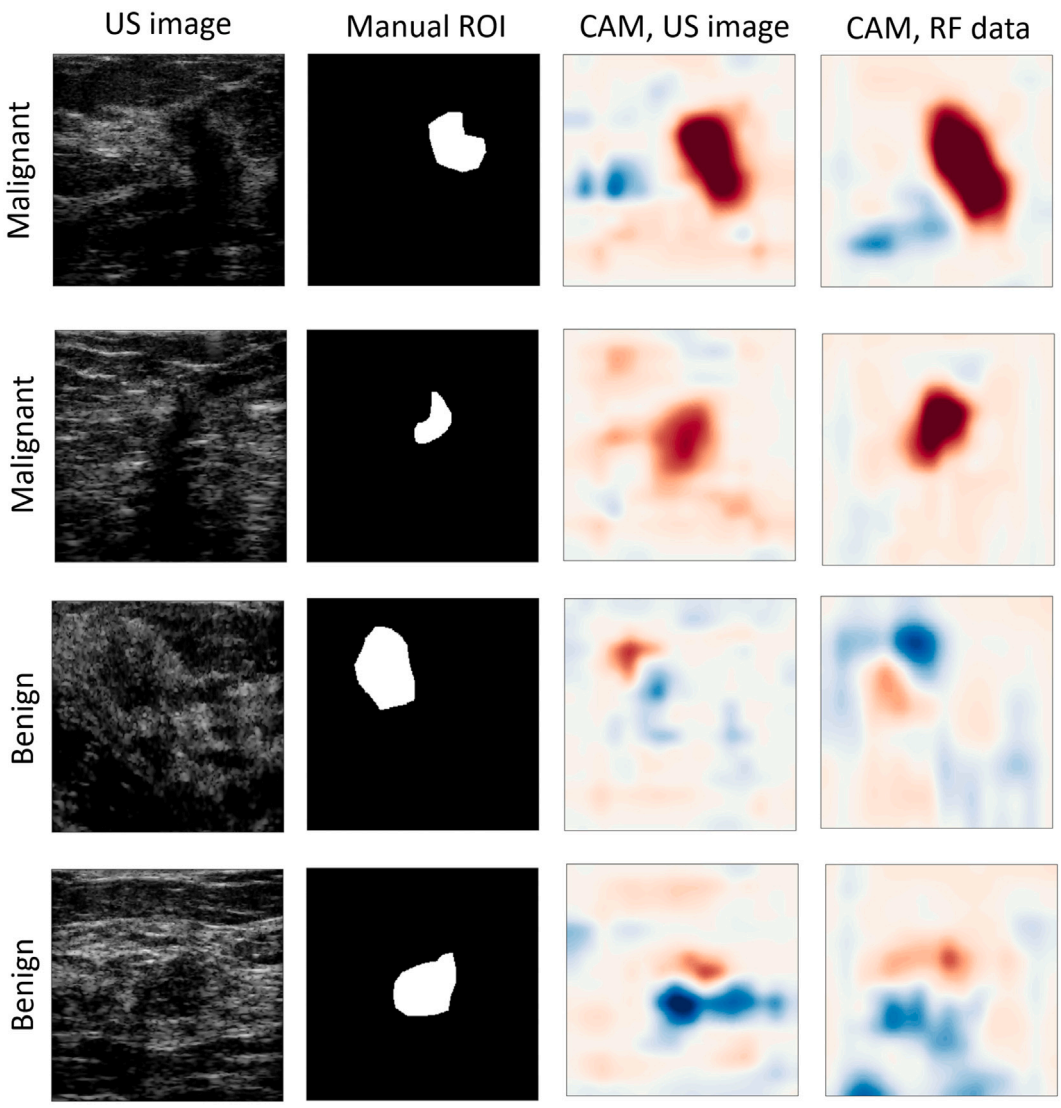

Fig. 9. Class activation maps (CAMs) obtained for the cases classified with high confidence by both networks. In this case, the *a posteriori* probability of malignancy was above 0.8 and below 0.2 for the malignant and benign masses, respectively. CAM generated for both networks and malignant masses presented strong positive activation (red color) in the mass areas. For the benign masses, the mass regions were occupied by lower positive activation combined with negative activation (blue color).

strategies to develop deep learning models, and commonly achieved good performance with AUC values above 0.85. For example, Han et al. developed a deep learning network based on a set of 7408 US images and achieved high AUC value of 0.96 [27]. Qi et al. trained and evaluated a network using a set of over 8145 US images and achieved high AUC value of 0.98 [32]. Additionally, CAMs calculated by the authors presented that the network highlighted mass areas to conduct classification decisions. Transfer learning techniques were investigated in several studies to devise deep models for breast mass classification. Antropova et al. used features extracted from pre-trained networks to train support vector machine classifiers [26]. The authors achieved AUC value of 0.9, based on a set of 2393 US images. In our previous work, we developed a deep transfer model based on a set of 882 US images and evaluated in on US images reconstructed using RF data from the OASBUD, the model achieved AUC value of 0.881 [29]. The network proposed in our study achieved AUC value of 0.874, which is a similar result. Nevertheless, we obtained a much lower score than the deep models proposed in previous studies, where much larger datasets of US images were used for training [26,27,32].

As far as we know, we used the RF data for the first time to address the problem of mass segmentation. Automatic segmentations calculated by the RF based network were compared with the manual segmentations prepared based on US images. Generally, Y-Nets achieved good segmentation scores, with the mean Dice scores of 0.64 and 0.60 for the US image based network and the RF data based model, respectively. While the US image based model achieved higher detection rate, the performance of the networks in respect to median Dice scores was similar and equal to around 0.746. However, while the biopsy served as the reference for the classification, segmentation was assessed based on manual segmentations provided by a single radiologist. Moreover, since the RF data cannot be used directly to outline breast masses, the manual segmentations were prepared based on US images, which could favor the US image based network. This issue raises a question how to assess the segmentations models trained with data that require processing to be visualized. Since there are no RF based deep learning models for the segmentation, we can only compare the performance of the US images model with the previous studies. Byra et al. developed a modified U-Net model and evaluated it on several publicly available datasets, including the US images from OASBUD [41]. Segmentation model trained without transfer learning on a set of 632 breast mass US images achieved median Dice score of 0.783 and detection rate of 0.78, performance similar to the reported for our method. However, due to the additional fine-tuning with OASBUD US images, the authors

achieved median Dice score and detection rate of 0.837 and 0.860, respectively. When it comes to the segmentation models based on U-Net, as the method in our paper, the following results were obtained in previous papers. Yap et al. developed a U-Net breast mass detection model based on two datasets and reported true positive rate of 0.87 (we averaged the results for both datasets). In another paper, Yap et al. utilized U-Net for breast mass segmentation and achieved mean Dice scores of 0.763 and 0.548 in the case of the malignant and benign masses [38]. Gomez-Flores et al. investigated the usefulness of several pre-trained CNNs for breast mass segmentation [52]. The U-Net model developed by the authors on a set of over 3000 US images achieved median intersection over union score of 0.804, translating to high Dice score of around 0.891.

To the best of our knowledge, there are no studies addressing the problem of the interpretability of deep learning models trained based on RF data. In the past, neural networks were commonly purely perceived as "black-box" models, but in the last few years various methods have been proposed to help understand how the neural networks work [53,54]. The results presented in our study show that the first block of the RF based Y-Net utilized convolutional filters presenting different oscillating patterns. Additional analysis revealed that the mean frequencies of the filters matched the frequency band of the imaging pulse of our research scanner. However, it remains to be investigated whether the filters extracted information from RF data that could be related to specific physical properties of tissues. Another visualization tool we used was the CAM method, which highlighted areas in input data important for the classification decision. We found that both deep learning models provided similar activation maps when it comes to cases classified with high confidence. This result suggests that the network conducts decisions based on characteristics of RF data extracted from breast mass region and its surroundings.

There are several issues with our study. First of all, the results presented in our study were obtained based on a relatively small set of RF data. The main issue with the RF data is that they are difficult to acquire. Standard US images can be collected as a part of routine clinical protocols, but the acquisition of RF data requires a research US scanner. Moreover, RF data, unlike US images, are not stored in hospital databases as a part of patient records. Therefore, large sets of RF data are difficult to assemble. In our study, the difference in performance between the RF data and the US image based models was small and insignificant statistically, suggesting that the US images can be efficiently used for the development of deep learning models when the RF data are not accessible. Presumably, the process of the B-mode US image reconstruction from the RF data did not remove the information required for the efficient breast mass diagnosis. Second issue is related to the robustness of the RF based network. Generally, QUS methods are considered to be more robust than the US image based methods, which can only provide approximate estimates of tissue physical properties, such as backscattering or attenuation. QUS methods may be used to assess physical properties for a large range of imaging frequencies. However, presented weights of the filters from the first convolutional block of the RF based network suggest that our approach may not be as robust as the QUS techniques. Since the mean frequencies of the filters matched the frequency band of the imaging pulse, modification of the imaging pulse frequency could result in extraction of worse performing features by the first convolutional filters and consequently undermine the processing of the data in deeper layers leading to worse classification and segmentation. This issue, however, remains to be studied. Third issue is that we did not assess the activation maps in a quantitative way. Ideally, a radiologist should review the CAMs and confront them with the established medical knowledge on characteristic US image features related to malignant and benign breast masses. For example, it would be interesting to take into account features listed in the Breast Imaging Reporting and Data System (BI-RADS) lexicon. This would provide more insight into the decision process conducted by the deep learning models. Classification decisions provided by networks, but supported by chaotic and noisy activation maps, should probably be interpreted with care. Our results on model interpretability presented in this work should be considered as preliminary.

Our future work will benefit from the further acquisition of RF data. We plan to investigate the usefulness of other US data types for the development of deep models. In this work, we compared RF data and US images based models, but it would be also feasible to include maps of QUS parameters or RF data spectrograms. The main aim of our work was to present the feasibility of using RF data for the deep learning based breast mass classification and segmentation. We did not focus on the engineering and evaluation of different deep learning models for these tasks. Our Y-Net model was based on the U-Net architecture, but several improvements of the U-Net have been proposed recently, such as the U-Net++ or attention-gated U-Net, which might be worth studying [55,56]. Moreover, in this work we trained the networks from scratch, but it would be also interesting to investigate the usefulness of various pre-trained models for the RF data processing [18]. For example, we could pre-train a network with computer vision data or simulated US data, and subsequently adjust it to process breast mass RF data.

6. Conclusions

In this work, we proposed a deep learning based approach for joint classification and segmentation of breast masses in US imaging. The proposed method utilized RF data to effectively address these two challenging tasks. Our results indicated that the RF data based network provided similar performance to the model based on the B-mode US images. We provided several interesting insights about the interpretability of the investigated networks. The methods proposed in this study, when fully developed, have the potential to help the radiologists with breast mass characterization in US.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Science Center of Poland (grant numbers 2014/13/B/ST7/01271, 2019/35/B/ST7/03792).

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 68 (6) (2018) 394–424.
[2] G.-G. Wu, L.-Q. Zhou, J.-W. Xu, J.-Y. Wang, Q. Wei, Y.-B. Deng, X.-W. Cui, C.F. Dietrich, Artificial intelligence in breast ultrasound, World J. Radiol. 11 (2) (2019) 19.
[3] J. Mamou, M.L. Oelze, Quantitative Ultrasound in Soft Tissues, Springer, 2013.
[4] L.X. Yao, J.A. Zagzebski, E.L. Madsen, Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors, Ultrason. Imaging 12 (1) (1990) 58–70.
[5] M.L. Oelze, J. Mamou, Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 63 (2) (2016) 336–351.
[6] P.M. Shankar, Ultrasonic tissue characterization using a generalized Nakagami model, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 48 (6) (2001) 1716–1720.
[7] P.-H. Tsui, C.-K. Yeh, C.-C. Chang, Y.-Y. Liao, Classification of breast masses by ultrasonic Nakagami imaging: a feasibility study, Phys. Med. Biol. 53 (21) (2008) 6027.
[8] A. Larrue, J.A. Noble, Modeling of errors in nakagami imaging: illustration on breast mass characterization, Ultrasound Med. Biol. 40 (5) (2014) 917–930.
[9] I. Trop, F. Destrempes, M. El Khoury, A. Robidoux, L. Gaboury, L. Allard, B. Chayer, G. Cloutier, The added value of statistical modeling of backscatter properties in the management of breast lesions at US, Radiology 275 (3) (2015) 666–674.

[10] M. Byra, A. Nowicki, H. Wróblewska-Piotrzkowska, K. Dobruch-Sobczak, Classification of breast lesions using segmented quantitative ultrasound maps of homodyned k distribution parameters, Med. Phys. 43 (10) (2016) 5561–5569.

[11] Z. Zhou, A. Gao, W. Wu, D.-I. Tai, J.-H. Tseng, S. Wu, P.-H. Tsui, Parameter estimation of the homodyned k distribution based on an artificial neural network for ultrasound tissue characterization, Ultrasonics 111 (2020) 106308.

[12] N. Uniyal, H. Eskandari, P. Abolmaesumi, S. Sojoudi, P. Gordon, L. Warren, R.N. Rohling, S.E. Salcudean, M. Moradi, Ultrasound RF time series for classification of breast lesions, IEEE Trans. Med. Imaging 34 (2) (2015) 652–661.

[13] P. Jarosik, Z. Klimonda, M. Lewandowski, M. Byra, Breast lesion classification based on ultrasonic radio-frequency signals using convolutional neural networks, Biocybern. Biomed. Eng. (2020).

[14] A. Han, M. Byra, E. Heba, M.P. Andre, J.W. Erdman Jr., R. Loomba, C.B. Sirlin, W.D. O'Brien Jr., Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks, Radiology (2020) 191160.

[15] S.J. Sanabria, J. Dahl, A. Pirmoazen, A. Kamaya, A. ElKaffas, Learning steatosis staging with two-dimensional convolutional neural networks: comparison of accuracy of clinical B-mode with a co-registered spectrogram representation of RF data, in: 2020 IEEE International Ultrasonics Symposium (IUS), 2020, pp. 1–4, http://dx.doi.org/10.1109/IUS46767.2020.9251329.

[16] T.N. Nguyen, A.S. Podkowa, T.H. Park, R.J. Miller, M.N. Do, M.L. Oelze, Use of a convolutional neural network and quantitative ultrasound for diagnosis of fatty liver, Ultrasound Med. Biol. (2020).

[17] A.K. Tehrani, M. Amiri, H. Rivaz, Real-time and high quality ultrasound elastography using convolutional neural network by incorporating analytic signal, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 2075–2078.

[18] A.K. Tehrani, H. Rivaz, Displacement estimation in ultrasound elastography using pyramidal convolutional neural network, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (12) (2020) 2629–2639.

[19] A.K. Tehrani, H. Rivaz, MPWC-Net++: evolution of optical flow pyramidal convolutional neural network for ultrasound elastography, in: Medical Imaging 2021: Ultrasonic Imaging and Tomography, Vol. 11602, International Society for Optics and Photonics, 2021, 1160206.

[20] D.Y. Chan, D.C. Morris, T.J. Polascik, M.L. Palmeri, K.R. Nightingale, Deep convolutional neural networks for displacement estimation in ARFI imaging, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2021).

[21] W.G. Flores, W.C. de Albuquerque Pereira, A.F.C. Infantosi, Improving classification performance of breast lesions on ultrasonography, Pattern Recognit. 48 (4) (2015) 1125–1136.

[22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[23] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J.G. Elmore, L. Shapiro, Y-Net: joint segmentation and classification for diagnosis of breast biopsy images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 893–901.

[24] H. Ryu, S.Y. Shin, J.Y. Lee, K.M. Lee, H.-J. Kang, J. Yi, Joint segmentation and classification of hepatic lesions in ultrasound images using deep learning, Eur. Radiol. (2021) 1–10.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 248–255.

[26] N. Antropova, B.Q. Huynh, M.L. Giger, A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets, Med. Phys. 44 (10) (2017) 5162–5171.

[27] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, Y.-K. Seong, A deep learning framework for supporting the classification of breast lesions in ultrasound images, Phys. Med. Biol. 62 (19) (2017) 7714.

[28] M. Byra, Discriminant analysis of neural style representations for breast lesion classification in ultrasound, Biocybern. Biomed. Eng. 38 (3) (2018) 684–690.

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

[30] J. Virmani, R. Agarwal, et al., Deep feature extraction and classification of breast ultrasound images, Multimedia Tools Appl. 79 (37) (2020) 27257–27292.

[31] M. Byra, Breast mass classification with transfer learning based on scaling of deep representations, Biomed. Signal Process. Control 69 (2021) 102828.

[32] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, Med. Image Anal. 52 (2019) 185–198.

[33] W.K. Moon, Y.-W. Lee, H.-H. Ke, S.H. Lee, C.-S. Huang, R.-F. Chang, Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks, Comput. Methods Programs Biomed. 190 (2020) 105361.

[34] E. Zhang, S. Seiler, M. Chen, W. Lu, X. Gu, BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis, Phys. Med. Biol. 65 (12) (2020) 125005.

[35] Z. Cao, G. Yang, Q. Chen, X. Chen, F. Lv, Breast tumor classification through learning from noisy labeled ultrasound images, Med. Phys. 47 (3) (2020) 1048–1057.

[36] M. Byra, T. Sznajder, D. Korzinek, H. Piotrzkowska-Wroblewska, K. Dobruch-Sobczak, A. Nowicki, K. Marasek, Impact of ultrasound image reconstruction method on breast lesion classification with deep learning, in: Iberian Conference on Pattern Recognition and Image Analysis, Springer, 2019, pp. 41–52.

[37] M.H. Yap, M. Goyal, F.M. Osman, R. Martí, E. Denton, A. Juette, R. Zwiggelaar, Breast ultrasound lesions recognition: end-to-end deep learning approaches, J. Med. Imaging 6 (1) (2018) 1–8.

[38] M.H. Yap, M. Goyal, F.M. Osman, R. Martí, E. Denton, A. Juette, R. Zwiggelaar, Breast ultrasound lesions recognition: end-to-end deep learning approaches, J. Med. Imaging 6 (1) (2018) 011007.

[39] M.H. Yap, M. Goyal, F. Osman, R. Marti, E. Denton, A. Juette, R. Zwiggelaar, Breast ultrasound region of interest detection and lesion localisation, Artif. Intell. Med. 107 (2020) 101880.

[40] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, C. Chang, Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model, Med. Phys. 46 (1) (2019) 215–228.

[41] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network, Biomed. Signal Process. Control 61 (2020) 102027.

[42] L. Han, Y. Huang, H. Dou, S. Wang, S. Ahamad, H. Luo, Q. Liu, J. Fan, J. Zhang, Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network, Comput. Methods Programs Biomed. (2019) 105275.

[43] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.

[44] S. Jadon, A survey of loss functions for semantic segmentation, in: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE, 2020, pp. 1–7.

[45] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[46] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.

[47] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics (1988) 837–845.

[48] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning, in: 12th {$USENIX$} Symposium on Operating Systems Design and Implementation ({$OSDI$} 16), 2016, pp. 265–283.

[49] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.

[50] M. Byra, A. Han, A.S. Boehringer, Y.N. Zhang, W.D. O'Brien Jr., J.W. Erdman Jr., R. Loomba, C.B. Sirlin, M. Andre, Liver fat assessment in multiview sonography using transfer learning with convolutional neural networks, J. Ultrasound Med. 41 (1) (2022) 175–184.

[51] H. Piotrzkowska-Wróblewska, K. Dobruch-Sobczak, M. Byra, A. Nowicki, Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions, Med. Phys. 44 (11) (2017) 6105–6109.

[52] W. Gómez-Flores, W.C. de Albuquerque Pereira, A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound, Comput. Biol. Med. 126 (2020) 104036.

[53] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. 73 (2018) 1–15.

[54] D.T. Huff, A.J. Weisman, R. Jeraj, Interpretation and visualization techniques for deep learning models in medical imaging, Phys. Med. Biol. 66 (4) (2021) 04TR01, http://dx.doi.org/10.1088/1361-6560/abcd17.

[55] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 3–11.

[56] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207.