Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Article

Breast lesion classification based on ultrasonic radio-frequency signals using convolutional neural networks

Piotr Jarosik [a,*], *Ziemowit Klimonda* [b], *Marcin Lewandowski* [c], *Michal Byra* [b]

[a] *Department of Information and Computational Science, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, Warsaw, Poland*

[b] *Department of Ultrasound, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, Warsaw, Poland*

[c] *Laboratory of Professional Electronics, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, Warsaw, Poland*

ARTICLE INFO

Article history:
Received 28 January 2020
Received in revised form
27 March 2020
Accepted 7 April 2020
Available online 00 Month 2020

Keywords:
Breast lesion classification
Convolutional neural networks
Deep learning
Radio-frequency signals
Ultrasound imaging

ABSTRACT

We propose a novel approach to breast mass classification based on deep learning models that utilize raw radio-frequency (RF) ultrasound (US) signals. US images, typically displayed by US scanners and used to develop computer-aided diagnosis systems, are reconstructed using raw RF data. However, information related to physical properties of tissues present in RF signals is partially lost due to the irreversible compression necessary to make raw data readable to the human eye. To utilize the information present in raw US data, we develop deep learning models that can automatically process small 2D patches of RF signals and their amplitude samples. We compare our approach with classification method based on the Nakagami parameter, a widely used quantitative US technique utilizing RF data amplitude samples. Our better performing deep learning model, trained using RF signals and their envelope samples, achieved good classification performance, with the area under the receiver attaining operating characteristic curve (AUC) and balanced accuracy of 0.772 and 0.710, respectively. The proposed method significantly outperformed the Nakagami parameter-based classifier, which achieved AUC and accuracy of 0.64 and 0.611, respectively. The developed deep learning models were used to generate parametric maps illustrating the level of mass malignancy. Our study presents the feasibility of using RF data for the development of deep learning breast mass classification models.

* *Corresponding author at*: Department of Information and Computational Science, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, Warsaw, Poland.
E-mail address: pjarosik@ippt.pan.pl (P. Jarosik).

https://doi.org/10.1016/j.bbe.2020.04.002

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women and the leading cause of cancer-related deaths worldwide [1]. Early detection and accurate diagnosis of breast cancer is crucial for the efficient treatment and survival of patients. Ultrasound (US) imaging is a popular medical imaging modality used for breast lesion characterization. US is low-cost, widely available and can be used to differentiate breast masses with high accuracy. However, US scans need to be collected by a skilled physician adept at operating a US scanner and locating the mass within the examined breast. Moreover, the assessment of breast masses in US images is subjective and depends on the radiologist's experience.

Various computer-aided diagnosis (CAD) systems have been developed to aid radiologists to objectively assess breast masses in US to improve classification performance and avoid unnecessary biopsies [2–5]. CAD systems commonly utilize US images to differentiate malignant and benign lesions. However, the appearance of tissues in US images depends on US scanner settings (e.g. scanner gain, compression). For instance, the texture of US images and visibility of object edges can be modified by using various image reconstruction methods and image post-processing algorithms [6]. Such modifications may have negative impact on the classification performance of the CAD systems utilizing US images [7]. Moreover, during the US image reconstruction process the information related to tissue structure carried by raw radio-frequency (RF) signals is partially lost due to irreversible compression, which is necessary to make the raw US data readable to the human eye. As a remedy, quantitative US (QUS) methods have been developed to extract features directly from RF data. QUS techniques commonly utilize small 2D patches of RF data to estimate features related to various local physical properties of tissue [8]. For instance, local spatial distribution of tissue micro-structures can be assessed based on RF signal amplitudes [9]. Commonly, the Nakagami and homodyned K distributions are used to model the statistics of RF signal amplitude. In the case of breast mass classification, shape parameters of the Nakagami and homodyned K distributions have been used to differentiate malignant and benign breast masses in several papers [10–13]. While QUS techniques are based on specific tissue models, Uniyal et al. developed a breast mass classification method based on generic hand-crafted features extracted from small 2D patches of RF data and US images [14]. In this case, various spectral features were calculated based on RF data. Moreover, features related to gray-level co-occurance matrix were estimated based on US images. Further, Ouyang et al. assessed the relative size of cancer tissue US scatters with the H-scan method to differentiate malignant and benign breast masses [15].

Deep learning methods based on convolutional neural networks (CNNs) are gaining importance in the medical image analysis field. CNN is an artificial neural network that includes at least one layer performing convolution-like operations [16]. CNNs have been frequently used to process time-domain [17] and spatial data [18]. They are considered a highly successful alternative to fully-connected neural networks, primarily due to the properties of convolutional layers such as translational equivariance, parameter sharing and sparse connectivity [16]. In US imaging, deep CNNs have been successfully applied to breast mass classification in several studies [19–25]. Due to relatively small data sets, authors of the aforementioned studies applied different transfer learning techniques with deep CNNs pre-trained on the ImageNet data set to develop the classifiers based on US images [26]. For instance, in [19] the authors used features extracted from the VGG19 CNN to train support vector machine classifiers. Meanwhile, in [20,25] pre-trained CNNs were fine-tuned to differentiate malignant and benign breast masses.

Deep learning models for breast mass characterization have also been developed for other imaging modalities, such as histology or mammography. Bardou et al. used CNNs for breast histology image classification, and achieved high accuracy in the differentiation of malignant and benign masses [27]. Mullooly et al. utilized deep learning methods to relate breast histology image features with mammographic breast density [28], which is a popular method to assess long-term risk of the breast cancer [29,30]. Diniz et al. used CNNs to classify breast tissue into dense and non-dense based in mammography images [31]. Al-masni et al. utilized a deep learning model named YOLO (You Only Look Once) to simultaneously detect lesions and assess malignancy of breast masses in mammography [32].

In this work, we propose a novel deep learning based approach to breast mass classification. Based on RF data collected from malignant and benign breast masses, we have developed deep CNNs that can automatically process RF signals and differentiate breast masses. As in the case of the previous studies on QUS techniques [10–13], our method is designed to process small 2D patches of RF data to locally characterize breast tissue. In contrast to QUS techniques, our approach is automatic and does not require any feature engineering or physical modelling – the deep learning model learns how to extract useful features for breast mass classification directly from the RF data. To the best of our knowledge, this is the first work on using deep learning methods to classify breast masses based on RF data. We investigate the usefulness of neural network models consisting of 1D and 2D convolutional layers developed using raw RF signals and their amplitudes. Moreover, we compare our approach with the classification method based on the Nakagami distribution.

2. Methods

2.1. Data set

To perform the experiments, we used the publicly available OASBUD data set, which includes US RF signal frames recorded from 52 malignant and 48 benign lesions [33]. Malignant masses were assessed with biopsy, while benign masses were assessed either with biopsy or a 2-year follow-up. For each mass, two orthogonal scans (transverse and longitudinal) were acquired using the Ultrasonix SonixTouch Research US scanner (Ultrasonix Inc., Canada) equipped with the L14-5/38 linear array transducer operating at a nominal central transmit frequency of 10 MHz. In each case, single focusing

beamforming was applied with the focal zone set at lesion depth. For each RF signal frame, 512 scan lines were collected, with the sampling frequency equal to 40 MHz. US images were reconstructed based on RF data, and used by an experienced radiologist to outline regions of interest (ROIs) indicating each breast mass area. More information about the data set can be found in the original paper [33]. Fig. 1 shows a reconstructed US image of a breast mass and a single RF signal corresponding to one of the scan lines.

Following the approaches from previous studies on QUS techniques, we developed the deep learning models using 2D patches of raw US data [10–13]. Separately, we extracted 2D patches of RF signals and their amplitude samples. The amplitude samples were calculated based on RF signals using Hilbert transform. The sliding window technique with stride equal to 1 mm was applied to divide each mass ROI into smaller sub-regions of 2×2 mm (104×27 samples). The window dimension corresponded approximately to three times the imaging pulse length and proved to provide good results in the case of Nakagami imaging [9]. In total, we obtained 3344 and 6301 2D data patches from the malignant and benign breast masses, respectively. Our approach to RF data analysis is presented in Fig. 2.

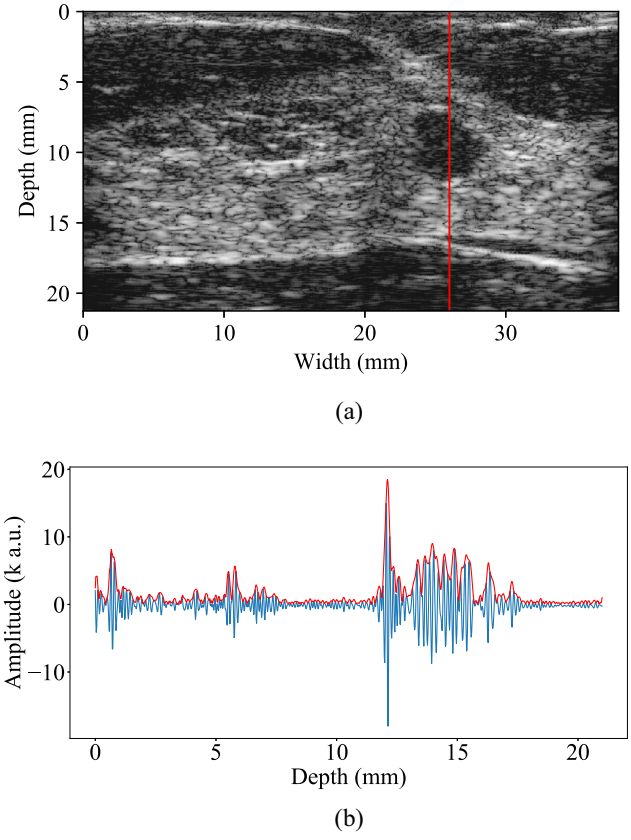


Fig. 1 – (a) US image of a breast mass reconstructed using RF data. The red line indicates one of RF signals scan lines. b) The indicated RF signal line (blue) and its envelope (red).

2.2. *Network architecture*

Let X be the set of all possible 2D RF data patches extracted from the malignant and benign breast masses. In general, any classifier can be expressed as a function $p(x;\theta)$, which assigns appropriate likelihood of malignancy to each patch $x \in X$. Here, θ stands for a set of model parameters to be determined during training. In this work we consider neural network architectures that are composed of convolutional, pooling and dense (*fully connected*) layers. A convolutional layer convolves input data x with trainable filters. For a 2-D case the output of the kth convolutional filter can be expressed in the following way [34]:

$$\phi(x, i, j)_k = (x * \theta_k)(i, j) + b_k, \tag{1}$$

where (i, j) is the pixel position in the kth feature map $\phi(x, i, j)_k$, x denotes input data, $*$ is the convolution operation and θ_k and b_k stand for the filter weights and bias term associated with the kth filter, both to be determined during the training procedure.

The pooling layer performs a down-sampling of input by computing a summary statistic over small parts of the data. In this work we compute the maximum and average over regions consisting of a couple data points (see Tables 1 and 2). We also use global average pooling (GAP), which computes the average over an entire feature map (the output of each convolutional filter separately). Pooling layers reduce the volume of processed data and thus may reduce the number of parameters of the further layers (e.g. dense layers). Also, a pooling layer is invariant to the (potentially small) translations of the input data.

The dense layer linearly transforms input by multiplying it by a matrix of trainable parameters θ:

$$\phi_l(x;\theta)_i = \sum_n \theta_{inl} x_n + b_{il}. \tag{2}$$

Dense layers are frequently used to reduce the dimensionality of input data; in particular, they can be used to reduce a vector of (latent) features to a single output of the neural network.

Frequently, a non-linear activation function is applied after each convolutional or dense layer; in this work we use a rectifier linear unit (ReLU) activation function for the output of convolutional layers [35]:

$$r(x) = \max(0, x) \tag{3}$$

and a sigmoid activation function for the dense layers:

$$s(x) = (1 + e^{-x})^{-1}. \tag{4}$$

Our approach to model development was inspired by several studies on the usage of deep neural networks for the processing of raw acoustic waveforms [36–38] and EEG signal decoding [39]. We investigated three different approaches to the binary classification of the 2D RF data patches.

The first network, denoted CNN-1D, was based on 1D convolutional layers and was developed to process RF signals. This model can be described as a composition $p(x) = f(\phi_s(\phi_t(x)))$,

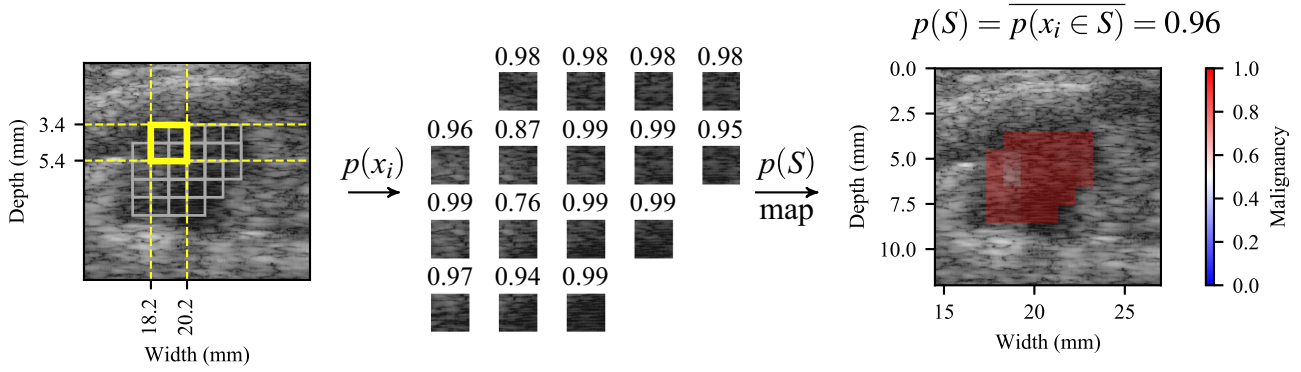


Fig. 2 – Classification procedure applied in this work. The sliding window technique (with stride equal to 1 mm) was used to extract 2 × 2 mm patches of RF data from breast mass ROIs. For each US data patch, the likelihood of malignancy $p(x_i)$ was computed with a CNN. The final decision on whether the lesion was malignant or benign was made on the basis of average likelihood $p(x_i)$, computed from all 2D patches obtained from all available examinations made for a given subject's lesion S. In addition, a map illustrating mass malignancy was generated in order to visualize the predictions made by the deep learning model.

where $\phi_t(x)$ consists of 1D convolution-pooling blocks processing each scan line separately in the longitudinal direction, and $\phi_s(x)$ consists of convolutional-pooling blocks that process the RF data patches in the transverse direction to extract the information about the relationships between scan lines. After the convolutional-pooling blocks, a global max pooling layer was used, followed by a dense layer with a sigmoid activation function applied to estimate the likelihood of malignancy. The architecture details of the proposed CNN-1D model are summarized in Table 1. The purpose of $\phi_t(x)$ was to extract temporal information present in each scan line separately. The usage of 1D convolutions reduced the impact of spatial information available in 2D patches by constraining the feature extraction process performed by $\phi_t(x)$ only to the time domain. This solution was inspired by the studies on EEG signal decoding with deep learning [39], where CNNs were developed to process EEG signals separately and then combine the information extracted from different EEG sensors for further processing. Moreover, processing each RF signal separately is a common approach to the estimation of QUS parameters, such as the backscatter and attenuation coefficients [40]. To take into account the potential relationship between features extracted from neighboring scan lines, we applied function $\phi_s(x)$, followed by a global max pooling layer to extract information about the presence of each feature in, at least, one scan line of the input data patch.

The second network architecture, denoted CNN-2D, was trained based on the envelope of RF signals. This model was designed to assess the usefulness of the spatial information present in RF signal amplitudes. The model contained three 2D convolution-max-pooling blocks, followed by a global average pooling layer and a dense layer with the sigmoid activation function. The architecture details are presented in Table 2. The CNN-2D model was inspired by the QUS techniques based on the Nakagami and homodyned K distributions, which extract information present in the envelope of RF signals. Similar deep learning models have been used for infant brain MR image segmentation [41].

The third network architecture, denoted CNN-1D-2D, was a combination of the two previous models (Fig. 3). In this case, we combined the CNN-1D and CNN-2D models by concatenating the outputs of the last global pooling layers, which were provided as the input to the dense layer followed by the sigmoid activation function. The aim of this solution was to train the model to simultaneously analyse RF signals and their amplitude samples.

We used the rectifier linear units (ReLU) activation functions to process outputs of the convolutional blocks [35] in all deep learning models. Moreover, we applied batch normalization after each convolutional block [42].

Table 1 – CNN-1D architecture.

Layer type	Description
Convolutional	16 filters 13 × 1
Average-pooling	Pooling and stride 2 × 1
Convolutional	16 filters 5 × 1
Max-pooling	Pooling and stride 2 × 1
Convolutional	32 filters 5 × 1
Max-pooling	Pooling and stride 2 × 1
Convolutional	32 filters 13 × 1
Convolutional	32 filters 1 × 6
Max-pooling	Pooling and stride 1 × 2
Convolutional	32 filters 1 × 6
Global max pooling (GMP)	
Dense	One output unit

Table 2 – CNN-2D architecture.

Layer type	Description
Convolutional	32 filters 9 × 4
Max-pooling	Pooling and stride 4 × 1
Convolutional	64 filters 5 × 5
Max-pooling	Pooling and stride 2 × 2
Convolutional	128 filters 5 × 5
Global average pooling (GAP)	
Dense	One output unit

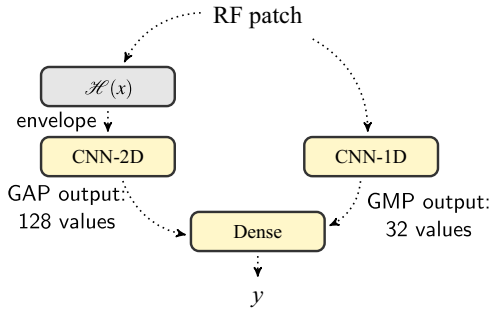


Fig. 3 – CNN-1D-2D architecture. The classification is performed using 2D patches of raw RF signals and their envelope samples computed with the 1-D Hilbert transform $H(\mathbf{x})$. Output vectors from the CNN-2D's Global Average Pooling layer (GAP) and CNN-1D's Global Max Pooling (GMP) layer were concatenated and provided as an input to a fully connected layer.

2.3. *Nakagami imaging*

The deep learning models proposed in this paper were compared with the Nakagami parameter based breast mass classifier [10,43]. For each 2D patch of RF signal amplitudes, we estimated the Nakagami parameter to assess local tissue scattering properties related to the spatial distribution of tissue microstructures. The probability density function of the Nakagami distribution can be described in the following way [44]:

$$f_N(A) = \frac{2m^m A^{2m-1}}{\Gamma(m)\Omega^m}\exp\left(\frac{m}{\Omega}A^2\right) \quad (5)$$

where A stands for the RF signal amplitude, $\Gamma(\cdot)$ is the Gamma function, m is the Nakagami parameter and Ω is the mean amplitude intensity. The Nakagami parameter is related to the spatial distribution of tissue microstructures (cell clusters) within the resolution cell of the imaging transducer. For $0.5 < m < 1$ the resolution cell contains a small number of scatterers, for $m = 1$ the resolution cell contains a large number of randomly distributed scatterers, for $m > 1$ the resolution cell is expected to include randomly distributed scatterers and regularly spaced scatterers. The maximum likelihood estimator was applied to calculate the Nakagami parameter [45].

2.4. *Model development and evaluation*

The data set was randomly divided 10 times into train, validation and test sets based on subject IDs. In each case, we used 50% of data for training, 20% for validation and 30% for testing. To improve the training of the models, we investigated two approaches to the normalization of input data [46]. The signals were standardized using heuristics transformations given by the following equations:

$$x_i^{\sim} = \frac{x_i - \mu}{\sigma} \quad (6)$$

and

$$x_i^{\sim} = \frac{x_i - \mu_i}{\sigma_i}, \quad (7)$$

where (μ, σ) are the mean and standard deviation of the amplitude samples in the whole training data set X_{train}, and (μ_i, σ_i) are the mean and standard deviation of the amplitude samples in a single 2D patch x_i. Signal pre-processing based on Eq. (6) maintains the information about the relative scale of signal value (mean intensity) between the patches; while for the second transformation, the amplitude of samples corresponding to each patch have the same standard deviation. We trained and evaluated the CNN-2D model using data transformed with both methods.

Deep learning models were developed using the training set to minimize the weighted negative cross entropy loss function. To address the problem of class imbalance, we adjusted the loss function inversely proportional to the class frequency in the training set. Weights of the models were initialized as random values drawn from a uniform distribution with limits determined using Glorot's method [47]. The stochastic gradient descent algorithm with initial learning rate equal to 10^{-4} and momentum 0.9 was used to perform back-propagation. We used a validation data set for an early stopping (100 epochs patience, maximum 500 epochs) procedure and to reduce the learning rate when validation loss reached plateau. We assessed the model's performance on a test data set using the area under the receiver operating curve (AUC) and balanced accuracy (arithmetic average of recall and specificity). We assigned malignant class to each RF data patch whenever the *a posteriori* probability of malignancy was higher than 0.5 ($p(x) > 0.5$); otherwise the patch was classified as benign. We calculated the classification performance metrics for two tasks. First, we evaluated how well the models performed in 2D patch classification. Second, we assessed the models in subject classification. In this case, we averaged the predictions corresponding to all patches extracted from a particular mass. This was performed to assess how accurately the models predicted the malignancy of breast mass. Similarly, the Nakagami parameters calculated for each 2D patch were averaged to determine the mean Nakagami parameter for the entire breast mass. Based on the mean Nakagami parameters calculated for each subject we developed a logistic regression classifier for the breast mass differentiation. The Welch's t-test was applied to compare the AUC values obtained for the better performing deep learning model and the Nakagami parameter based classifier.

>Based on the patches extracted with the sliding window technique, we used the deep learning models to generate parametric maps illustrating the likelihood of malignancy for each breast mass. The likelihood of malignancy obtained for each sub ROI was overlaid on the US image. All calculations performed in this study were done in Python 3.6. The networks were implemented in Keras 2.2.4 with the Tensorflow 1.12 backend [48,49].

3. Results

3.1. *Impact of the RF signal normalization method*

We trained and evaluated the CNN-2D model using inputs normalized as in Eqs. (6) and (7); the results are presented in

Table 3 – Performance scores obtained by the CNN-2D model using two different RF data normalization schemes.

	AUC (±std. dev)	Accuracy (±std. dev.)
Classification per subject		
std. (6)	0.776 (±0.075)	0.691 (±0.073)
std. (7)	0.762 (±0.072)	0.691 (±0.085)
classification per 2D patch		
std. (6)	0.775 (±0.075)	0.710 (±0.062)
std. (7)	0.746 (±0.072)	0.682 (±0.059)

Table 4 – Classification performance obtained by the deep learning models and the Nakagami parameter based classifier.

	AUC (±std. dev.)	Accuracy (±std. dev.)
classification per subject		
CNN-2D	0.762 (±0.072)	0.691 (±0.085)
CNN-1D	0.730 (±0.104)	0.646 (±0.072)
CNN-1D-2D	0.772 (±0.083)	0.701 (±0.073)
Nakagami parameter	0.640 (±0.093)	0.611 (±0.051)
classification per 2D patch		
CNN-2D	0.746 (±0.072)	0.682 (±0.059)
CNN-1D	0.648 (±0.045)	0.602 (±0.032)
CNN-1D-2D	0.721 (±0.067)	0.653 (±0.056)
Nakagami parameter	0.624 (±0.069)	0.593 (±0.051)

Table 3. We achieved slightly better classification performance for the first standardization, both for classification per patch and per subject. Hence, for the remaining experiments we decided to employ the second standardization scheme.

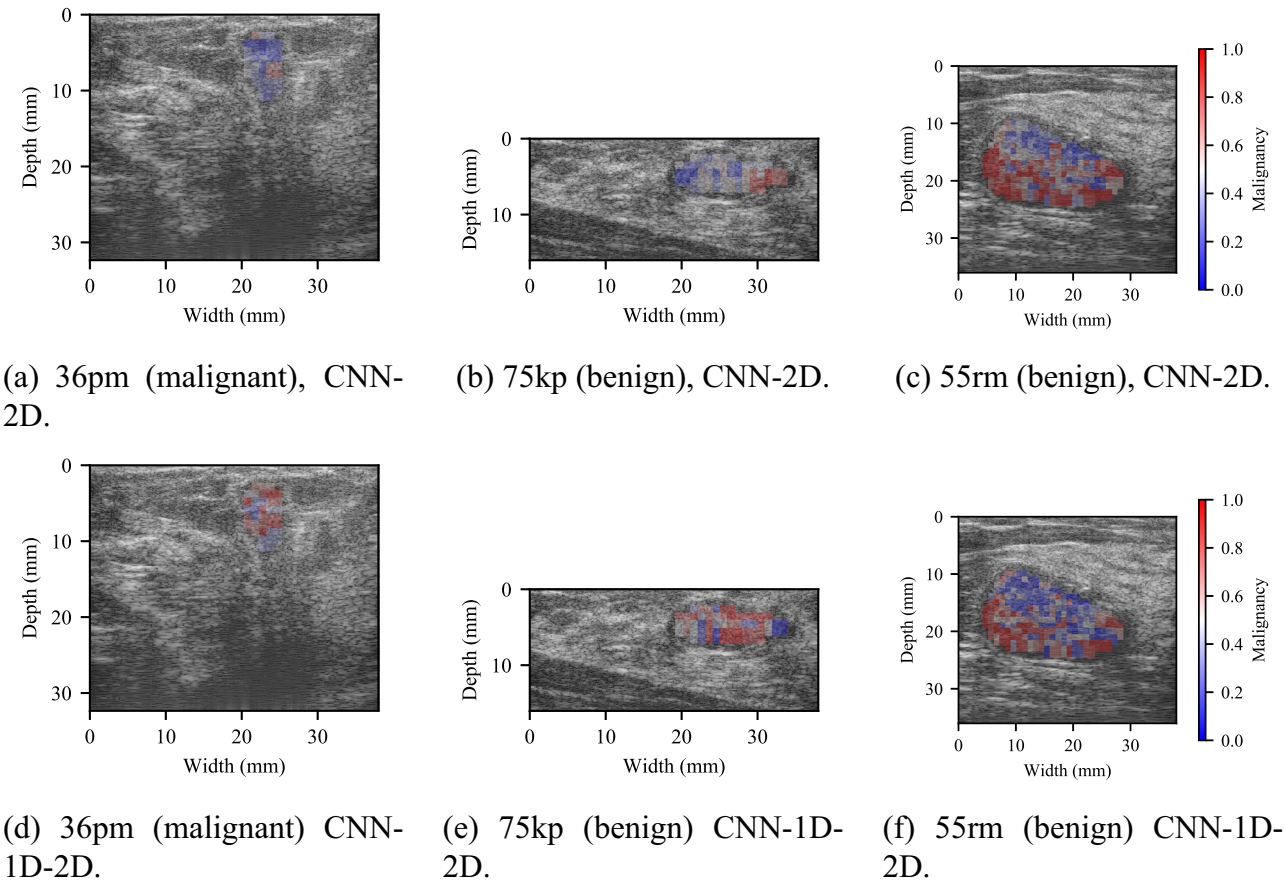
3.2. *Classification performance*

Table 4 presents the classification performance achieved by the deep learning models and the Nakagami parameter based classifier. The best average results for the subject classification task were obtained by the CNN-1D-2D model, with the AUC and accuracy values equal to 0.772 and 0.701, respectively. As opposed to the CNN-2D model, incorporating the CNN-1D model improved the classification performance by a small margin. The Nakagami parameter based classifier achieved an AUC value of 0.64 in the subject classification task. This result was significantly lower than the AUC value obtained for the CNN-1D-2D model (p-value <0.05).

In the patch classification task, the model based on the RF signal envelope achieved the best results, with AUC and accuracy values of 0.746 and 0.682, respectively. In comparison, the CNN-1D model developed using RF data achieved a much lower AUC value of 0.648. To further compare the results obtained for the patch classification task, we examined the number of misclassified 2D patches in one of the test sets for each subject individually. These are presented in Table 5. In comparison to the CNN-2D model, the usage of the CNN-1D-2D model improved the accuracy for subjects with the highest

Table 5 – Classification error computed for each training subject separately.

Id.	Class	# ROIs	CNN-2D error	CNN-1D-2D error
36pm	malignant	37	34 (91.9%)	17 (45.9%)
30nh_1	benign	19	15 (78.9%)	9 (47.4%)
114hd	malignant	22	17 (77.3%)	10 (45.5%)
176JP	malignant	4	3 (75.0%)	2 (50.0%)
55rm	benign	392	261 (66.6%)	207 (52.8%)
65gs	benign	13	7 (53.8%)	2 (15.4%)
14mw	malignant	23	12 (52.2%)	8 (34.8%)
188	malignant	22	11 (50.0%)	8 (36.4%)
51ag	benign	96	43 (44.8%)	54 (56.2%)
10jw	malignant	114	45 (39.5%)	50 (43.9%)
102ks	malignant	378	142 (37.6%)	93 (24.6%)
115r	malignant	35	13 (37.1%)	19 (54.3%)
75kp	benign	75	27 (36.0%)	56 (74.7%)
111kp	benign	84	25 (29.8%)	27 (32.1%)
110jb	malignant	191	52 (27.2%)	76 (39.8%)
47sm	benign	98	26 (26.5%)	27 (27.6%)
23wp	benign	125	32 (25.6%)	49 (39.2%)
30nh_2	benign	73	17 (23.3%)	20 (27.4%)
41ba	malignant	199	38 (19.1%)	56 (28.1%)
31nh	benign	64	8 (12.5%)	18 (28.1%)
180IM	malignant	663	70 (10.6%)	127 (19.2%)
112bw	benign	33	3 (9.1%)	4 (12.1%)
25ci	benign	69	6 (8.7%)	16 (23.2%)
181JO	malignant	307	18 (5.9%)	29 (9.4%)
205AT	malignant	106	1 (0.9%)	0 (0.0%)
49ez	benign	7	0 (0.0%)	2 (28.6%)
46sm	benign	2	0 (0.0%)	0 (0.0%)
42sw	benign	2	0 (0.0%)	1 (50.0%)
226MP	malignant	45	0 (0.0%)	1 (2.2%)
200AT	malignant	56	0 (0.0%)	4 (7.1%)



(a) 36pm (malignant), CNN-2D. (b) 75kp (benign), CNN-2D. (c) 55rm (benign), CNN-2D.

(d) 36pm (malignant) CNN-1D-2D. (e) 75kp (benign) CNN-1D-2D. (f) 55rm (benign) CNN-1D-2D.

Fig. 4 – Malignancy maps developed for subjects with ID number 36pm, 75kp, 55rm using the CNN-2D (first row) and CNN-1D-2D (second row) models. For subjects 36pm, 55rm we observed an improvement in patch classification accuracy; for subject 75kp we noticed a decrease in performance.

number of incorrectly classified patches (e.g. subjects: 36pm, 30nh_1, 55rm); however, it also resulted in several misclassifications (e.g. subjects: 51ag, 75kp, 41ba). CNN-1D-2D decreased the number of subjects that had at least 50% of incorrectly classified patches: from 8 for CNN-2D (subjects: 36pm, 30nh_1, 114hd, 176JP, 55rm, 65gs, 14mw, 188) to 6 for CNN-1D-2D (subjects: 75kp, 51ag, 115r, 55rm, 42sw, 176JP). It also decreased the number of subjects with all patches correctly classified (0% error) – from 5 to 2. The highest error rate was 91.9% (subject 36pm) for CNN-2D and 74.7% for CNN-1D-2D (subject 75kp).

Fig. 4 presents parametric maps illustrating the likelihood of mass malignancy obtained for the cases classified with different levels of confidence. The maps were generated using the CNN-1D-2D model.

4. Discussion

Our study confirmed the feasibility of using CNNs for breast mass classification based on patches of raw RF US data. The highest AUC value, equal to 0.772, was obtained for the CNN-1D-2D model. The approach proposed in this study was most similar to that developed by Uniyal et al., where the classification method based on generic features extracted from RF signals achieved an AUC value of 0.68 on a set of 22 breast lesions [14]. However, in contrast to this work, we did not apply any feature engineering to develop the classifier. Instead, we trained the CNN model to automatically process RF data and perform 2D patch classification. Our better performing deep learning model achieved significantly higher scores than the Nakagami parameter based classifier. This result suggests that deep learning based analysis of RF signals may provide better results than standard QUS techniques. However, the obtained results for the Nakagami parameter were worse in comparison to those reported in previous studies on breast mass classification. For instance, Larrue and Noble used the Nakagami distribution to model RF signal amplitudes and achieved an AUC value of 0.78 on a set of 37 breast lesions [12]. Liao et al. proposed a similar model based on the Nakagami distribution, and obtained an AUC value of 0.84 based on a set of 130 breast lesions [10]. Shankar et al. combined a Nakagami parameter computed from two ultrasound images and achieved an AUC value of 0.8316 [44]. The superior performance of Nakagami parameter based classifiers reported in previous studies could result from the fact that the authors did not apply any cross-validation to calculate the performance scores, possibly rendering the obtained results overly optimistic. The discrepancy in results could also be caused by the differences in data sets and imaging procedures.

While our approach achieved better classification performance in comparison to the Nakagami based model, the obtained AUC values were significantly lower than those reported for deep learning breast mass classification systems utilizing US images [19,20,24,25]. These models were developed using transfer learning with pre-trained deep neural networks, and achieved AUC values above 0.85 in breast mass classification. For instance, Byra et al. developed a deep learning model using US images reconstructed based on the RF data from the OASBUD data set (the data used in our study), and achieved an AUC of 0.819 [24]. This AUC value was higher than that obtained using our better performing model, 0.772. The superior performance of the deep learning models developed using US images may be partially explained by the fact that the malignancy of breast masses is related to features that are not present in small patches of RF data extracted from breast mass area. For instance, morphological features of mass contour play an important role in breast mass differentiation [3]. Virmani et al. showed that a hybrid CAD system that uses a combination of morphological features extracted from a despeckled US image and texture features extracted from the unprocessed US image can achieve 96% accuracy [7]; similarly, high performance has been reported for the segmentation task [5]. We trained our models using RF data only, and thus information about mass appearance in respect to surrounding tissues was not utilized. Moreover, transfer learning was not applied in our study, as the deep learning models were trained from scratch. The aforementioned methods utilized deep learning models trained on large sets of natural images. As far as we know, there are no open access data sets of RF data that could be used to pre-train a deep learning model. In our study, the models were trained and evaluated using a relatively small data set of 200 RF data frames collected from 100 breast masses. Supposedly, the classification performance of the deep learning models should increase with data volume.

The approach proposed in this study has several advantages. First, it can be used to create parametric maps illustrating the likelihood of malignancy within the examined breast mass. Second, while our method was developed using small patches of RF data, handcrafted features or patient characteristics can be easily incorporated into the model. For instance, morphological features could be combined with the features produced by the penultimate layer of the network. This way, the model would be trained to simultaneously take into account mass morphology and to extract information from RF data. Moreover, this idea is not limited to handcrafted features: QUS parametric maps, elastograms and US images can also be incorporated into the model to provide more information about breast mass physical properties [50,51]. Nevertheless, there are several issues related to our work which we would like to address in the future. First, we have so far only investigated two approaches to model development, but it would be interesting to explore other possibilities. For instance, the neural networks could be trained using spectrograms calculated based on RF signals. It might also be productive to investigate the usefulness of other deep learning architectures, for instance recurrent neural networks. Second, we trained our models from scratch, but it might be feasible to apply transfer learning. US simulation programs, such as Field II, could be used to artificially generate large volumes of RF data from tissue mimicking numerical phantoms [52]. We could then use the artificial RF data to initially pre-train the deep learning model, and fine-tune it using RF data collected from human tissues.

5. Conclusions

In this work, we developed a breast mass classification method based on convolutional neural networks and ultrasound radio-frequency data. Our results confirmed the feasibility of developing models that can process small 2D patches of radio-frequency data and differentiate between malignant and benign breast masses. Our better performing deep learning model outperformed the Nakagami parameter based classifier. As opposed to other quantitative ultrasound techniques, our method can automatically process RF data to extract features useful for the classification. The proposed approach to tissue characterization is general and can be applied to processing RF data collected from other human tissues.

Author statement

Piotr Jarosik – Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing.

Ziemowit Klimonda – Conceptualization, Data Curation, Writing – Original Draft, Writing – Review & Editing.

Marcin Lewandowski – Conceptualization, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision.

Michal Byra – Conceptualization, Methodology, Data Curation, Writing – Original Draft, Writing – Review & Editing, Supervision.

Conflict of interest

The authors do not have any conflicts of interest.

Acknowledgement

This work has not received any kind of funding.

REFERENCES

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J Clin 2018;68:394–424.

[2] Cheng H-D, Shan J, Ju W, Guo Y, Zhang L. Automated breast cancer detection and classification using ultrasound images: a survey. Pattern Recognit 2010;43:299–317.

[3] Flores WG, de Albuquerque Pereira WC, Infantosi AFC. Improving classification performance of breast lesions on ultrasonography. Pattern Recognit 2015;48:1125–36.

[4] Wu G-G, Zhou L-Q, Xu J-W, Wang J-Y, Wei Q, Deng Y-B, et al. Artificial intelligence in breast ultrasound. World J Radiol 2019;11:19.
[5] Virmani J, Agarwal R, et al. Assessment of despeckle filtering algorithms for segmentation of breast tumours from ultrasound images. Biocybern Biomed Eng 2019;39:100–21.
[6] Yu X, Guo Y, Huang S-M, Li M-L, Lee W-N. Beamforming effects on generalized nakagami imaging. Phys Med Biol 2015;60:7513.
[7] Virmani J, Agarwal R, et al. Effect of despeckle filtering on classification of breast tumors using ultrasound images. Biocybern Biomed Eng 2019;39:536–60.
[8] Oelze ML, Mamou J. Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63:336–51.
[9] Tsui P-H, Chang C-C. Imaging local scatterer concentrations by the nakagami statistical model. Ultrasound Med Biol 2007;33:608–19.
[10] Liao Y-Y, Tsui P-H, Li C-H, Chang K-J, Kuo W-H, Chang C-C, et al. Classification of scattering media within benign and malignant breast tumors based on ultrasound texture-feature-based and nakagami-parameter images. Med Phys 2011;38:2198–207.
[11] Trop I, Destrempes F, El Khoury M, Robidoux A, Gaboury L, Allard L, et al. The added value of statistical modeling of backscatter properties in the management of breast lesions at us. Radiology 2014;275:666–74.
[12] Larrue A, Noble JA. Modeling of errors in nakagami imaging: illustration on breast mass characterization. Ultrasound Med Biol 2014;40:917–30.
[13] Byra M, Nowicki A, Wróblewska-Piotrzkowska H, Dobruch-Sobczak K. Classification of breast lesions using segmented quantitative ultrasound maps of homodyned k distribution parameters. Med Phys 2016;43:5561–9.
[14] Uniyal N, Eskandari H, Abolmaesumi P, Sojoudi S, Gordon P, Warren L, et al. Ultrasound rf time series for classification of breast lesions. IEEE Trans Med Imaging 2015;34:652–61.
[15] Ouyang Y, Tsui P-H, Wu S, Wu W, Zhou Z. Classification of benign and malignant breast tumors using h-scan ultrasound imaging. Diagnostics 2019;9:182.
[16] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
[17] Lang KJ, Waibel AH, Hinton GE. A time-delay neural network architecture for isolated word recognition. Neural Netw 1990;3:23–43.
[18] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1:541–51.
[19] Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 2017;44:5162–71.
[20] Han S, Kang H-K, Jeong J-Y, Park M-H, Kim W, Bang W-C, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys Med Biol 2017;62:7714.
[21] Byra M. Discriminant analysis of neural style representations for breast lesion classification in ultrasound. Biocybern Biomed Eng 2018;38:684–90.
[22] Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform 2018;22:1218–26.
[23] Yap MH, Goyal M, Osman FM, Martí R, Denton E, Juette A, et al. Breast ultrasound lesions recognition: end-to-end deep learning approaches. J Med Imaging 2018;6:011007.
[24] Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. Med Phys 2019;46:746–55.
[25] Qi X, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q, et al. Automated diagnosis of breast ultrasonography images using deep neural networks. Med Image Anal 2019;52:185–98.
[26] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis (IJCV) 2015;115:211–52. http://dx.doi.org/10.1007/s11263-015-0816-y
[27] Bardou D, Zhang K, Ahmad SM. Classification of breast cancer based on histology images using convolutional neural networks. IEEE Access 2018;6:24680–93.
[28] Mullooly M, Bejnordi BE, Pfeiffer RM, Fan S, Palakal M, Hada M, et al. Application of convolutional neural networks to breast biopsies to delineate tissue correlates of mammographic breast density. NPJ Breast Cancer 2019;5:1–11.
[29] Destounis SV, Santacroce A, Arieno A. Update on breast density, risk estimation, and supplemental screening. Am J Roentgenol 2020;214:296–305.
[30] Rebolj M, Blyuss O, Chia KS, Duffy SW. Long-term excess risk of breast cancer after a single breast density measurement. Eur J Cancer 2019;117:41–7.
[31] Diniz JOB, Diniz PHB, Valente TLA, Silva AC, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed 2018;156:191–207.
[32] Al-Masni MA, Al-Antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. Comput Methods Programs Biomed 2018;157:85–94.
[33] Piotrzkowska-Wróblewska H, Dobruch-Sobczak K, Byra M, Nowicki A. Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions. Med Phys 2017;44:6105–9.
[34] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–324.
[35] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics; 2011. p. 315–23.
[36] Hoshen Y, Weiss RJ, Wilson KW. Speech acoustic modeling from raw multichannel waveforms. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2015. pp. 4624–8.
[37] Golik P, Tüske Z, Schlüter R, Ney H. Convolutional neural networks for acoustic modeling of raw time signal in lvcsr. Sixteenth Annual Conference of the International Speech Communication Association; 2015.
[38] Sainath TN, Weiss RJ, Senior A, Wilson KW, Vinyals O. Learning the speech front-end with raw waveform cldnns. Sixteenth Annual Conference of the International Speech Communication Association; 2015.
[39] Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, et al. Deep learning with convolutional neural networks for eeg decoding and visualization. Hum Brain Mapp 2017;38:5391–420.
[40] Nam K, Zagzebski JA, Hall TJ. Quantitative assessment of in vivo breast masses using ultrasound attenuation and backscatter. Ultrason Imaging 2013;35:146–61.
[41] Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage 2015;108:214–24.
[42] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift; 2015, arXiv:1502.03167 (arXiv preprint).

[43] Tsui P-H, Yeh C-K, Chang C-C, Liao Y-Y. Classification of breast masses by ultrasonic nakagami imaging: a feasibility study. Phys Med Biol 2008;53:6027.
[44] Shankar PM. A general statistical model for ultrasonic backscattering from tissues. IEEE Trans Ultrason Ferroelectr Freq Control 2000;47:727–36.
[45] Lin J-J, Cheng J-Y, Huang L-F, Lin Y-H, Wan Y-L, Tsui P-H. Detecting changes in ultrasound backscattered statistics by using nakagami parameters: comparisons of moment-based and maximum likelihood estimators. Ultrasonics 2017;77:133–43.
[46] LeCun YA, Bottou L, Orr GB, Müller K-R. Efficient backprop. Neural networks: tricks of the trade. Springer; 2012. p. 9–48.
[47] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; 2010. p. 249–56.
[48] Chollet F. Keras; 2018, https://github.com/fchollet/keras.
[49] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). 2016. pp. 265–83.
[50] Byra M, Piotrzkowska-Wróblewska H, Dobruch-Sobczak K, Nowicki A. Combining nakagami imaging and convolutional neural network for breast lesion classification. 2017 IEEE International Ultrasonics Symposium (IUS); 2017. pp. 1–4.
[51] Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics 2016;72:150–7.
[52] Jensen JA. Field: a program for simulating ultrasound systems. 10TH Nordicbaltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1; 1996. pp. 351–3.