Ultrasound radio-frequency time series for finding malignant breast lesions

Anonymous Author(s)
Affiliation
Address
email

Abstract

This project provides an insight into ultrasound-based solutions for breast lesion characterization to reduce the patient recall rate after mammography screening. In this work, ultrasound radio frequency time series analysis is performed for separating benign and malignant breast lesions with similar B-mode appearance. The radio frequency time series method is versatile and requires only a few seconds of imaging with no need for additional instrumentation. This study employs the spectral and fractal features of ultrasound radio frequency time series along with a machine learning framework with leave-one-patient-out cross validation of the classification. Support vector machines, bagged decision trees (random forest), and naive Bayes methods are used and compared in this context. For clinical relevance, cancer probability maps are also produced, by estimating the posterior malignancy probability of regions of size 1 mm^2 in the suspicious lesions. Recorded area under the receiver operating curve is, 0.79 using SVM, 0.74 using random forest, and 0.68 using naive Bayes classification. All classifiers successfully classified 6 out of 7 patients with malignant breast lesions and 4 out of 5 patients with benign lesions, with success defined as correct classification of at least 80% of the 1 mm^2 regions. The above findings suggests that ultrasound radio frequency time series along with the developed machine learning framework can help in differentiating malignant from benign breast lesions.

1 Introduction

In the United States, it is estimated that 226,870 (29%) of all new cancer cases among women will be breast cancer [1]. Breast cancer also ranks first in cancer related deaths among women of 20-59 years [1]. Mammography is routine for screening asymptomatic patients; however, it is not very effective in identifying benign and malignant breast lesions [2, 3]. The high rate of recall for biopsy after mammography emphasizes the need to augment the Breast Imaging-Reporting and Data System (BI-RADS). Breast ultrasound is used as a supplement to mammography in distinguishing benign from malignant in non-palpable breast lesions. Ultrasound has a high sensitivity, however rather low specificity. In some studies, a specificity of 31% was recorded [4]. Currently, almost 75% of the biopsies carried out after radiological diagnosis turn out to be benign [2, 5]. Thus, there exists a need for reducing the number of breast biopsies and improving ultrasound-based diagnosis.

To improve the performance of ultrasound-based techniques for breast cancer imaging, researchers have widely explored strain-based elastography for tumor classification with high success rates [6, 7, 8, 9, 10]. The average area under the Receiver Operating Characteristic (ROC) curve for ultrasound elastography was recorded as 0.90 [6], 0.85 [7], and 0.92 [8]. Freehand elastography depends on compression applied to the tissue by the probe. This compression is applied manually through the transducer [10], or by the use of a "shaker"[8].

Recently, a tissue typing method based on a time series of Radio Frequency (RF) ultrasound signals was proposed which could complement B-mode and elastography techniques for breast lesion classification. The advantage of this method is that it does not require a vibration mechanism. In the past, RF time series has been proven successful in detecting prostate cancer and differentiating animal tissue types [11, 12, 13]. The credibility of RF time series for tissue typing has been established through analysis and experiments. A model has been developed to relate the variations of the US backscattering to the variations in tissue temperature and sound speed that take place during the RF time series scanning procedures [14]. The measurements of the variations in the US backscattering are then used in a tissue classifier. In this work, for the first time, the performance of RF time series in separating benign and malignant breast lesions *in vivo*, is reported.

For the purpose of this project, a machine learning framework was developed for quantitative analysis of spectral and fractal parameters extracted from RF time series. In this study, three different classification algorithms were used to generate highly sensitive and specific malignancy maps that can be used for decision support in biopsy recommendation. The proposed approach can be part of the overall solution for multiparametric ultrasound imaging of breast cancer.

2 Methods

2.1 Data Collection

An RF time series is formed by the sequence of RF echoes received from one location in the tissue over time. To acquire the RF time series, one keeps the ultrasound probe and the tissue fixed in place and acquires frames of RF signals. In this method of analysis, the tissue typing parameters are extracted from the temporal changes of signal as opposed to the classical method of spectral analysis on spatial segments of the RF signals [15].

In order to study the utility of RF time series analysis in separating cancerous breast lesions from benign findings, a study approved by the Clinical Research Ethics Board at the University of British Columbia was devised. Patients referred to ultrasound-guided biopsy, based on mammography screening, were consented for data collection during biopsy for this study. The study was conducted between September 2012 and January 2013. Data was obtained on a Sonix Touch ultrasound machine (Ultrasonix Medical Corp., Richmond, BC, Canada). The research platform provided by the manufacturer enabled acquisition of raw RF signals in real time. For every subject, the sonographer first performed a preliminary ultrasound scan to find the suspected lesion. Once the lesion was located, the sonographer would hold her hand steady for 4 seconds while a computer program stored the RF data into the memory and consequently saved it in a file. Imaging was performed with an L14-5/38 ultrasound transducer at a center frequency of 10 MHz and a depth of 4 cm. Each RF line was sampled at 40 MHz and a total of 128 A-lines were acquired for each RF frame. With these image settings, a frame-rate of 98 frames per second was achieved. The data collection of each subject was followed by a routine ultrasound exam and a core needle biopsy of the lesion under ultrasound guidance by the physician. The first 12 cases are reported here. Biopsy results for these cases showed seven malignant lesions all of the invasive ductal carcinoma type and five benign cases all of the fibroadenoma type.

2.2 Features

Only the first 256 (out of 490) RF frames (2.6 sec) were used for calculating the features. This was done to minimize the effects of patient motion on the analysis. The biopsied lesions were divided into 1 mm^2 regions of interest (ROIs). The tissue typing features were extracted from these ROIs. In the RF domain, this ROI size was equivalent to 3×52 samples each forming a time series. Note that from the 12 lesions in the dataset, a total of 641 ROIs were extracted. Among these, 241 were malignant. In other words, the first 241 samples of the training data belonged to the malignant class and the rest 400 samples belonged to the benign class.

The following seven spectral and fractal parameters were extracted from each time series and averaged to form the feature vector describing an ROI.

Spectral parameters: The frequency spectrum was estimated by calculating the FFT-based periodogram of the Hamming windowed time series. This estimated spectrum was divided into four

frequency bands and each averaged to deliver a feature. In other words, the first four features (Features 1-4) were the average of the frequency spectrum in [0,π/4) , [π/4,π/2) , [π/2,3π/4) , [3π/4,π] frequency bands in the discrete frequency domain. Note that the sampling rate here is equivalent to the frame rate of the ultrasound machine. Two other spectral features were the intercept (Feature 5) and the slope (Feature 6) of a regression line fitted to the magnitude of the spectrum versus normalized frequency.

Fractal dimension: Feature 7 was the average fractal dimension of RF time series in a region of interest. In this context, fractal dimension (FD) is a measure of the nonlinear complexity of the signal. For calculation of the FD, the algorithm proposed by Higuchi [16] was used, which decomposes the signal into different scales and evaluates the signal complexity. Higuchi's algorithm was used with 16 levels of decomposition for the time series of length 256.

2.3 Classification and estimation of posterior malignancy probabilities

Support vector machine (SVM): SVM is a widely used maximum margin classifier. In the classification process using SVM, the data is first mapped to a significantly higher dimension and then the optimum hyperplane that separates the data into two classes is found. Training the data includes searching for the best hyperplane that maximizes the orthogonal distance between the datasets that are closest in the two classes and the decision boundary. The C++ SVM implementation called LIBSVM and reported in [17] was used for this purpose. The Radial Basis Function (RBF) kernel defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{1}$$

was employed and the parameter values of the RBF exponent and the slack variable weight coefficient were set with a grid search. SVM classifier can be used with other kernels like, linear, sigmoid, and polynomial however, for our problem RBF kernel worked better. For the SVM classification problem described in this report only two parameters had to be found, C, that governs the trade-off between the separating margin and the slack variable penalty, and γ, that is the RBF kernel parameter. The parameters C and γ were found by doing an exhaustive search between 1-1000 and performing a tenfold cross validation. Posterior class probabilities were calculated using Platt's algorithm [18] as follows. Assume that the SVM hyperplane obtained after training is $W\phi(F) + b$ where F is the feature vector and ϕ is the kernel function. A sigmoid function of form:

$$P_c = (cancer|(W\phi(F_i) + b)) = \frac{1}{1 + exp(A(W\phi(F_i) + b) + B)} \tag{2}$$

is calculated for mapping the test feature vector F_i to posterior malignancy probability P_c. Maximum likelihood estimation from the observations for which the true labels are known (training data) is fitted to calculate the values of the parameters A and B. P_c values are used as the threshold parameter for generating the ROC curves and malignancy maps.

Bagged decision trees (Random Forest): Random forest implementation (TreeBagger) within the MATLAB Statistics Toolbox was used. Bagged decision trees or random forest is a popular classification algorithm employed by many researchers [19]. Random forest consists of an ensemble of N decision trees, $\{T_1(F), ..., T_N(F)\}$, where $F = \{f_1, ..., f_n\}$ is an n dimensional feature vector containing observation features for an ROI. The trees in the ensemble produce outputs $\{\hat{Y}_1 = T_1(F), ..., \hat{Y}_N = T_N(F)\}$, where $\hat{Y}_i, i = 1, ..., N$ is the predicted label of an observation by the i_{th} tree based on a cutoff of 0.5 on the posterior class probability. The overall posterior class probability for the malignant class is obtained by simple averaging on all trees in the ensemble (forest). The final label prediction $\hat{Y}$ is done based on the overall posterior class probability. In this work, only 500 trees were grown and the trees were constructed shallow to avoid over-fitting the data. To increase the randomness of the classifier and avoid error due to noise only 3 features (out of 7) were randomly selected to determine the best split at each node of the tree. Also, for bootstrapping only 2/3 of the training data was used. The training algorithm implemented in MATLAB closely follows the method described by [19].

Naive Bayes classification: The implementation within MATLAB Statistics Toolbox was used. The kernel density estimation was used for the distribution of the features. Unlike SVM, Naive Bayes method is probabilistic in nature and the outcome is class probability as opposed to label. However, an underlying assumption in Bayesian approach as implemented is that features should be condi-

tionally independent. Later in the report it is shown that the naive Bayes classifier performed worst due to high correlation between the features (as shown in Figure 1).

Cross validation: In all classifiers, the cross-validation was leave-one-patient out in order to avoid the possible optimistic bias introduced by the batch structure of the data. The ultimate aim of the proposed classification method is to provide accurate predictions on future subjects and only a leave-one-patient-out scheme can test that ability (as opposed to k-fold cross-validation).

3 Results

A heat map showing the correlation between the features is given in Figure 1. A high degree of correlations is observed between spectral parameters. The feature selection search returned a relatively small subset of features that included features 2 and 4 and 7. The search was exhaustive and was performed on the SVM separately with the goal of maximizing the area under ROC curve with leave-one-patient-out cross validation. Exhaustive search was feasible due to the small number of features that resulted in only 127 non-empty subsets.

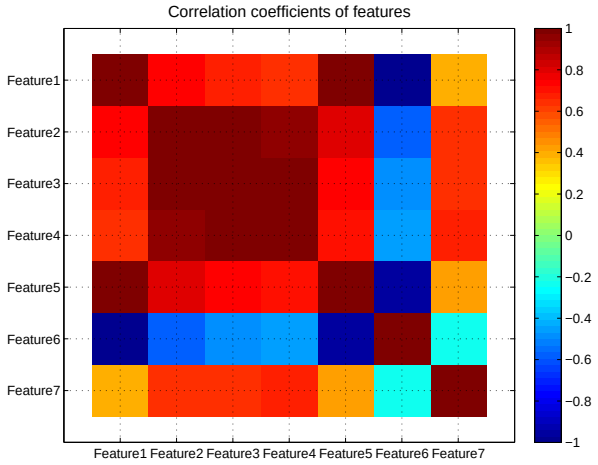

Figure 1: The correlation coefficients of features as a heatmap

The classification results were validated using the biopsy result of the patients. Table 1 is the summary of the histology reports (IDC stands for invasive ductal carcinoma and FA stands for fibroadenoma). The bulk result of the classification is also reported in this table. In six out of seven malignant cases, all three classifiers successfully predicted the label in leave-one-patient-out cross validation. It should be noted that success was defined as obtaining the correct label for 80% of the 1 mm^2 ROIs extracted from a lesions. For the one malignant case which was misclassified, further investigation of the B-mode image revealed calcification in parts of the area of the pathalogic finding. For benign cases, the classifiers were successful in four out of five cases. The result was consistent among the three classifiers, meaning that the same case was misclassified by all three classifiers.

The posterior class probabilities for the 641 ROIs of size 1 mm^2, generated by leave-one-patient-out training and testing scheme, was used to generate the ROC curves for the three classifiers. This resulted in an area under the curve of 0.79 using SVM, 0.74 using random forest, and 0.68 using naive Bayes. The curves are depicted in Figure 3. The weak performance of the naive Bayesian method could be attributed to the fact that the selected three features are not conditionally independent as observed in Figure 1.

Using a cutoff value of P_c=0.5 for the posterior class probabilities, Table 2 reports the sensitivity and specificity of the classifiers. A sensitivity of 86% was obtained using SVM. It is also notable that the methodology is specific with specificity values of 84.5% and 85% for SVM and random forest, respectively. SVM and random forest outperformed the naive Bayes classifier. SVM was most sensitive (86.7%) and random forest was most specific (86%).

Table 1: Histology report for 12 patients

Patient	BIRADS	Tissue Type	Biopsy Result	Classification Result
Patient 1	5	IDC	Malignant	Malignant
Patient 2	5	IDC	Malignant	Malignant
Patient 3	5	IDC	Malignant	Benign
Patient 4	4B	IDC	Malignant	Malignant
Patient 5	4C	IDC	Malignant	Malignant
Patient 6	4C	IDC	Malignant	Malignant
Patient 7	4B	IDC	Malignant	Malignant
Patient 8	3	FA	Benign	Benign
Patient 9	4A	FA	Benign	Benign
Patient 10	4A	FA	Benign	Malignant
Patient 11	4A	FA	Benign	Benign
Patient 12	4A	FA	Benign	Benign

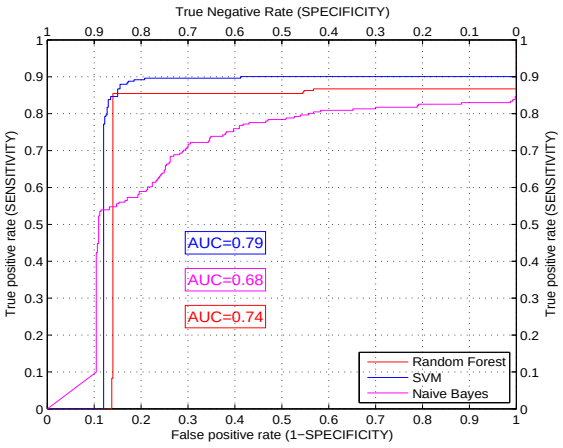

Figure 2: ROC curves for the three classifiers

As Figure 3 shows, by plotting the values of class probabilities on the area of the lesions, a map of malignancy is obtained. These maps are generated using the SVM classifier. This can be potentially used for guiding the biopsy in real time.

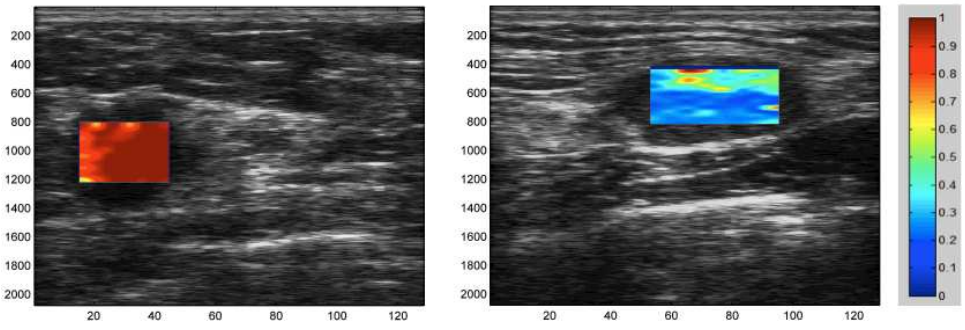

Figure 3: The malignancy map created by plotting the value of posterior malignancy probability (P_c) of the ROIs overlaid on the B-mode image. The image on the left is from a malignant case, and one on the right is a benign case. As expected, the left image shows high probability of cancer (reddish) and the right image shows a low probability of cancer (bluish)

Table 2: Sensitivity and specificity values for classifiers.

Classification Algorithm	Sensitivity	Specificity
SVM	86.7%	84.5%
Random Forest	85.4%	86%
Naive Bayes	70.5%	69.3%

4 Conclusion

In this work, the use of ultrasound RF time series as a method for detecting malignant breast lesions, is reported. This is a pressing clinical need and can potentially reduce the rate of recall for biopsy after mammography. The results of this study suggests that, by using a selection of spectral and fractal parameters extracted from temporal RF signals, within a machine learning classification framework, one can obtain high sensitivity and specificity for this problem. The present dataset includes mostly benign cases of BI-RADS 4. These are very likely to be sent to biopsy and as the results show, the biopsy could have been avoided with the use of this methodology in four out of five cases.

The time series method in this work required less than 3 seconds of data. It was found that by asking the patients to hold their breath, the amount of motion can be minimized to reduce the impact of motion on the analysis. Lastly, it can be concluded that RF time series is a practical and accurate method for sonographic augmentation of the BI-RADS criteria. Previous work has shown that a multiparametric ultrasound approach is potentially capable of improving the diagnostic value of imaging and it is argued here that RF time series can be a component of that approach.

References

[1] Siegel, R., Naishadham, D., Jemal, A.: Cancer statistics, 2012. CA: A Cancer Journal for Clinicians **62**(1) (2012) 10–29

[2] Brown, M., Houn, F.: Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. American Journal of . . . (1995)

[3] Giger, M.L., Al-Hallaq, H., Huo, Z., Moran, C., Wolverton, D.E., Chan, C.W., Zhong, W.: Computerized analysis of lesions in US images of the breast. Academic radiology **6**(11) (November 1999) 665–74

[4] Buchberger, W., Niehoff, A.: Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. Seminars in Ultrasound, . . . **4** (2000) 325–336

[5] Poplack, S., Carney, P., Weiss, J.: Screening Mammography: Costs and Use of Screening-related Services1. . . . (6) (2005) 79–85

[6] Burnside, E., Hall, T., Sommer, A.: Differentiating Benign from Malignant Solid Breast Masses with US Strain Imaging. Radiology (November) (2007) 401–410

[7] Chung, S.Y., Moon, W.K., Choi, J.W., Cho, N., Jang, M., Kim, K.G.: Differentiation of benign from malignant nonpalpable breast masses: a comparison of computer-assisted quantification and visual assessment of lesion stiffness with the use of sonographic elastography. Acta radiologica (Stockholm, Sweden : 1987) **51**(1) (February 2010) 9–14

[8] Itoh, A., Ueno, E., Tohno, E., Kamma, H.: Breast Disease: Clinical Application of US Elastography for Diagnosis. Radio (May) (2006) 341–350

[9] Hall, T.J., Sommer, a.M., Burnside, E.S., Sisney, G.a., Hesley, G.K., Hangiandreou, N.J., Svensson, W.E.: ROC Analysis of Ultrasound Elasticity Imaging of Breast Abnormalities. 2006 IEEE Ultrasonics Symposium (2006) 2048–2051

[10] Zhu, Q.L., Jiang, Y.X., Liu, J.B., Liu, H., Sun, Q., Dai, Q., Chen, X.: Real-time ultrasound elastography: its potential role in assessment of breast lesions. Ultrasound in medicine & biology **34**(8) (August 2008) 1232–8

[11] Moradi, M., Abolmaesumi, P., Siemens, D.R., Sauerbrei, E.E., Boag, A.H., Mousavi, P.: Augmenting detection of prostate cancer in transrectal ultrasound images using SVM and RF time series. IEEE transactions on bio-medical engineering **56**(9) (September 2009) 2214–24

[12] Moradi, M., Abolmaesumi, P., Mousavi, P.: Tissue typing using ultrasound RF time series: Experiments with animal tissue samples. Medical Physics **37**(8) (2010) 4401

[13] Imani, F., Daoud, M., Moradi, M., Abolmaesumi, P., Mousavi, P.: Tissue classification using depth-dependent ultrasound time series analysis: in-vitro animal study. **7968** (2011) 79680F–79680F–7

[14] Daoud, M.I., Mousavi, P., Imani, F., Rohling, R., Abolmaesumi, P.: Tissue classification using ultrasound-induced variations in acoustic backscattering features. IEEE Transactions on Biomedical Engineering **In Press** (2012)

[15] Feleppa, E.J., Kalisz, A., Sokil-melgar, J.B., Lizzi, F.L., Liu, T., Rosado, A.L., Shao, M.C., Fair, W.R., Wang, Y., Cookson, M.S., Reuter, V.E., Heston, W.D.W.: Typing of Prostate Tissue by Ultrasonic Spec trurn Analysis. **43**(4) (1996) 609–619

[16] Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena (1988)

[17] Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using the second order information for training SVM. Machine Learning Research **6**(1) (2005) 1889–1918

[18] Platt, J.C.: Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: Advances in Large Margin Classifier. MIT Press, Cambridge, MA (2000)

[19] Breiman, L.: Random forests. Machine learning (2001) 5–32