

Prostate Cancer Probability Maps Based on Ultrasound RF Time Series and SVM Classifiers

Mehdi Moradi¹, Parvin Mousavi¹, Robert Siemens², Eric Sauerbrei³,
Alexander Boag⁴, and Purang Abolmaesumi¹

¹ School of Computing

² Department of Urology

³ Department of Diagnostic Radiology

⁴ Department of Pathology and Molecular Medicine,

Queen's University, Kingston, ON, Canada

purang@cs.queensu.ca

Abstract. We describe a very efficient method based on ultrasound RF time series analysis and support vector machine classification for generating probabilistic prostate cancer colormaps to augment the biopsy process. To form the RF time series, we continuously record ultrasound RF echoes backscattered from tissue while the imaging probe and the tissue are stationary in position. In an *in-vitro* study involving 30 prostate specimens, we show that the features extracted from RF time series are significantly more accurate and sensitive compared to two other established categories of ultrasound-based tissue typing methods. The method results in an area under ROC curve of 0.95 in 10-fold cross-validation.

1 Introduction

Prostate cancer accounted for 29% of cancer cases among American men and terminated the lives of 31,350 North Americans in 2007 [1]. The common clinical diagnosis method for the disease is histopathologic analysis of biopsy samples acquired under ultrasound guidance. However, most prostate tumors lack visually distinct appearances on medical images. Therefore, pathologically significant cases of cancer can be missed during biopsy, resulting in false negative or repeated trials. The goal of our research is to augment the ultrasound images with tissue typing information that can be used for targeting during biopsy.

There is a wealth of literature on ultrasound-based diagnosis of prostate cancer [2]. Texture features of B-scan images and spectral features (Lizzi-Feleppa features [3]) extracted from calibrated average spectrum of RF signals have been used along with numerous classification approaches for tissue typing [4]. Also, elastography, an automatic method for measuring the elasticity of tissue, has shown promising results in diagnosis of the disease [5]. However, the aforementioned approaches have not gained wide-spread clinical acceptance due to a variety of reasons including limited accuracy and overheads such as the need for calibration or tissue compression imposed on the clinical routine.

Recently, a new paradigm in tissue typing has been proposed by our group [6,7]. The core of the idea is that if a specific location in tissue undergoes continuous interaction with ultrasound, the time series of echoes from that location would carry “tissue typing” information. In other words, although variations in the intensity of a spatial sample of RF echo over time are partly due to different sources of noise, they depend on the tissue type as well. We have used RF time series for tissue typing in animal tissues [7] and have also employed them along with neural network classifiers for detection of prostate cancer on a small size dataset [6]. This paper is the first report of a major clinical *in-vitro* study involving 30 patients in which the performance of RF time series features has been evaluated. We report quantitative comparisons between the RF time series features, and Lizzi-Feleppa (LF) and texture features. We describe a classification approach based on an extended version of Support Vector Machines (SVM) [8] that provides posterior class probabilities for classified regions of interest (ROIs). Using these probability values, we generate cancer probability maps. Finally, we study the effects of imaging depth and acoustic power on the tissue typing capabilities of RF time series.

The paper is organized as follows: Section 2 provides the details about our data collection, feature extraction and selection techniques, and SVM-based classification and color-map generation; Section 3 provides the classification results and colormaps; Section 4 analyzes the effects of different parameters on RF time series features; and, Section 5 summarizes and concludes the paper.

2 Method

2.1 Data Collection

We have performed an extensive process of ultrasound and histopathology data collection from patients who chose prostatectomy as their treatment option at a local hospital. Data was collected from 30 patients. Extracted prostate specimens were suspended in a water bath, and were scanned along transverse planes that were 4 *mm* apart. The ultrasound RF data was collected using a Sonix RP (Ultrasonix Inc., Richmond, BC, Canada) ultrasound machine. A transrectal probe (BPSL9-5/55/10) was used with the central frequency set to 6.6 *MHz*. To form the RF time series, we continuously acquired 112 frames of RF data, at the rate of 22 frames per second (fps) from each cross-section of the tissue. Therefore, the length of RF time series in our analysis was 112 for each tissue cross-section. The maximum imaging depth was set to 4.5 *cm*. The RF sampling rate was 20 *MHz* and the number of bits per RF sample was 16.

After ultrasound data acquisition, the prostate specimens were dissected along the scanned cross-sections. The first cross-section was marked with two parallel needles that were visible in ultrasound images as two lines and defined a plane. The consequent parallel cuts were made at 4 *mm* intervals. Histopathological analysis of whole mount slides were acquired and used as the gold standard. The process of registering the histopathology maps to the RF frames was performed

manually. Due to the elevation beam width of ultrasound signals and also inevitable errors caused by the low precision cutting process in the pathology lab, the accurate match between tumor boundaries in ultrasound and histopathology images was challenging. Therefore, only a limited number of ROIs from cross-sections for which a high level of confidence of registration could be achieved were selected. This process amounted to 1478 normal and 856 cancerous ROIs selected from 46 cross-sections. It should be noted that although only 46 of the cross-sections could be used in the training process, the developed methodology was used to generate cancer colormaps on all acquired cross-sections.

The ROI dimension was $1\text{ mm} \times 3.5\text{ mm}$. This included four segments of RF lines, each of length 96 samples. Our investigations show that RF time series features extracted from ROIs even as small as 1 mm^2 are effective for tissue typing. However, the frequency spectrum of the RF segments extracted for Lizzi-Feleppa method can not be evaluated in very short RF segments.

2.2 Features

RF time series features (S1-S6,FD). To create RF time series, we continuously recorded RF echo signals backscattered from tissue, while the imaging probe and the tissue were fixed in position. Samples of RF signals collected over time from a fixed spot of tissue formed one RF time series.

Spectral features of the RF time series (S1-S6): Each RF time series is a discrete signal of length 112 in our dataset. After removing the mean, and zero-padding of the length of the time series to 128, we used the FFT algorithm to estimate the frequency spectrum corresponding to each spatial RF sample in an ROI and then, averaged each component of the spectrum over the ROI. We normalized the average spectrum to values in [0-1] interval. The first four RF time series features (S1, S2, S3 and S4) were the average value of the normalized spectrum in four quarters of the frequency range. In other words, they summarized the low, mid-low, mid-high, and high frequency components of the signal, respectively. We also fit a regression line to values of the spectrum (versus normalized frequency). The intercept (S5) and the slope (S6) of this line were used as two more features [6].

Fractal dimension (FD) of RF time series: It has been shown that the fractal dimension of the RF time series computed based on Higuchi's algorithm [9] contains tissue typing information [7]. We computed the FD ($K_{max} = 16$) of all the time series within an ROI and averaged them to get one feature per ROI.

Lizzi-Feleppa features (LF1,LF2,LF3). Lizzi, Feleppa and their colleagues have shown that the intercept extrapolated to zero frequency (LF1), the average slope (LF2), and midband value (LF3) of a line fitted to the mid-band portion of the calibrated frequency spectrum of segments of RF A-lines, can be used as the signature of cancerous and normal tissue types in prostate [3]. We closely followed the methodology described in [3]. The calibration data for Lizzi-Feleppa method was acquired from the surface of a flat glass plate in a water bath at the transducer focal zone, with minimum amplifier gain and flat TGC. The values

of slope and midband value were corrected for an assumed linear attenuation coefficient of $0.5 \text{ dB/MHz} - \text{cm}$. Feleppa *et al.* have averaged the spectral frequency on 10 RF segments. However, in our study, increasing the number of RF line segments from 4 to 8 (which doubles the ROI size from 3.5 mm^2 to 7 mm^2) had no significant effect on the efficiency of LF features (compare the results in Table 1 for ROI sizes of 3.5 and 7 mm^2).

Texture features. We used 12 texture features from the B-scan equivalents of the collected RF frames [4]. They included four statistical moments of the pixel intensities (mean, std, skewness and kurtosis), and eight features from the co-occurrence matrices: correlation, energy, contrast, and homogeneity computed for co-occurrence distance of $l = 1$ and separately for 0° and 90° directions.

2.3 SVM Classifier and Posterior Class Probabilities

Support vector machines use a kernel function to map the input data to a higher dimension space where a hyperplane can separate the data into different classes. The process of training a SVM classifier is equivalent to finding this optimal hyperplane in a way that minimizes the error on the training dataset and maximizes the perpendicular distance between the decision boundary and the closest data points in classes [10]. In a two class case, if the training dataset consists of N feature vectors $\{x_1, \dots, x_N\}$ with class labels $y_i \in \{1, -1\}$, then the SVM training problem is equivalent to finding W and b such that [10]:

$$1/2 \times W^T W + C \sum_{i=1}^N \xi_i \quad (1)$$

is minimized subject to:

$$y_i(W^T \phi(x_i) + b) \geq 1 - \xi_i \quad , \quad i = 1, \dots, N \quad (2)$$

where $\xi_i \geq 0$ are the so-called slack variables that allow for misclassification of noisy data points, and parameter $C > 0$ controls the trade-off between the slack variable penalty and the margin [10]. The function $\phi(x)$ maps the data to a higher dimensional space. This new space is defined by its kernel function: $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. The above problem can be formulated as a quadratic optimization process. The details of the solution and its implementation can be found in [11]. We used the Gaussian Radial Basis Function (RBF) kernel:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (3)$$

This was firstly due to the fact that RBF kernel has only one parameter (γ) to adjust. Also, we found SVM classifiers based on RBF kernel more accurate than linear, sigmoid, and polynomial kernels in case of our problem.

The SVM is merely a decision machine: If $f(x_n) = W^T \phi(x_n) + b > 0$, then the class label for x_n is $y_n = 1$. In other words, SVM does not provide posterior class probabilities ($P(\text{class}|\text{input})$). In order to generate probability colormaps

and also Receiver Operating Characteristic (ROC) curves based on SVM classification, one needs posterior probabilities of normal and cancer classes:

$$P_{can}(x_n) = p(y = 1|f(x_n)) \quad (4)$$

where $P_{can}(x_n)$ stands for probability of x_n being cancerous and $P_{nor}(x_n) = 1 - P_{can}(x_n)$. Platt [12] has extended SVM for probability estimates by training an SVM and then finding the parameters of a sigmoid function of form:

$$P = (y_n = 1|f(x_n)) = \frac{1}{1 + \exp(Af(x_n) + B)} \quad (5)$$

to map the values of $f(x_n)$ to posterior probabilities. The values of parameters A and B are fit using a maximum likelihood estimation from the training set [12]. We used class probabilities generated with this method for creating the probabilistic cancer maps and also as decision thresholds for ROC curve building.

We used the publicly available C++ implementation of the SVM algorithms known as LIBSVM [11]. The entire dataset was normalized prior to training by setting the maximum value of each feature to 1 and the minimum to 0. For parameter selection (C and γ), we exhaustively searched the parameter space $1 \leq C \leq 100, 1 \leq \gamma \leq 100$ with steps of length 1. For each set of parameters, 10-fold cross-validation was performed: we trained the SVM using 90% of the data samples, classified the remaining 10%, and repeated the procedure for all 10 portions of the data. The computational expense of this exhaustive search was feasible on our parallel computing server (Sun Microsystems, Sun Fire V890).

2.4 Feature Selection

We separately performed exhaustive searches to find the best subset of texture, LF and time series features to maximize their tissue typing accuracy. These three best subsets were combined and subjected to an additional exhaustive search to find a feature vector to generate probabilistic cancer colormaps. This two stage method of feature selection was not guaranteed to provide the highest possible accuracy. An exhaustive search over all 22 features combined could have resulted in a better performing subset. However, it required examining $\sum_{n=0}^{22} C_n^{22} = 2^{22}$ subsets which was computationally infeasible. Feature space reduction based on Principle Component Analysis (PCA) was an alternative solution. We chose the two-stage method because: 1) the two-stage method enabled us to compare the performance of the three approaches of tissue typing; 2) the PCA-based features did not exceed the accuracy of the subset found in our two-stage search (regardless of the number of principle components included).

3 Results

Performance of different groups of features: We found the optimal parameters (C and γ) of the RBF-based SVM classifier separately for each group of features and used them during feature selection (see Table 1 for parameter values).

Exhaustive search for the best subset within each of the three categories of features resulted in a two-dimensional feature vector among texture features (Mean and correlation at 0°), a two-dimensional feature vector among Lizzi-Feleppa features (LF1,LF3), and a six-dimensional feature vector (S2,S3,S4,S5,S6,FD) among the proposed time series features. As Table 1 demonstrates, while texture and LF features show a similar performance, time series features significantly outperform both of them with accuracy of $82.5 \pm 1.6\%$ (compared to $74.4 \pm 2.7\%$ and $75.9 \pm 1.3\%$). More importantly, the time series approach is the only one to provide a high sensitivity. There was no advantage in increasing the size of ROIs from 3.5 to 7 mm^2 (compare the two right columns of Table 1).

To acquire the best combination of features for cancer probability colormaps, we combined the selections of features from the three groups into a 10-dimensional feature vector and performed the exhaustive search on them. The seven dimensional subset (Mean,S2,S4,S6,FD,LF1,LF3) was found as the optimal subset considering its performance and dimension. As the last row of Table 1 reports, an accuracy of 92.1% with sensitivity of 88.3% and specificity of 94.1% was achieved with this selection of features. The area under the ROC curve acquired using this subset of features with 10-fold cross-validation was 0.95.

Table 1. Comparison of feature groups on the corresponding optimal SVM classifiers

Feature Group	SVM parameters	results for 3.5 mm^2 ROI size	results for 7 mm^2 ROI size
Best subset of Texture features (Mean and correlation at 0°)	RBF	Acc: $74.4 \pm 2.7\%$	$75.4 \pm 3.4\%$
	$C = 50$	Sen: $49.0 \pm 5.1\%$	$51.7 \pm 7.1\%$
	$\gamma = 94$	Spe: $89.2 \pm 2.6\%$	$89.1 \pm 3.9\%$
Best subset of Lizzi-Feleppa features (LF1,LF3)	RBF	Acc: $75.9 \pm 1.3\%$	$72.3 \pm 2.8\%$
	$C = 50$	Sen: $57.7 \pm 3.8\%$	$52.8 \pm 6.3\%$
	$\gamma = 97$	Spe: $84.9 \pm 3.0\%$	$82.1 \pm 3.7\%$
Best subset of Time series features (S2,S3,S4,S5,S6,FD)	RBF	Acc: $82.5 \pm 1.6\%$	$79.7 \pm 3.4\%$
	$C = 37$	Sen: $79.6 \pm 2.4\%$	$75.5 \pm 5.2\%$
	$\gamma = 25$	Spe: $83.4 \pm 3.8\%$	$82.1 \pm 3.9\%$
Best subset of the combination of the three groups above: (Mean,S2,S4,S6,FD,LF1,LF3)	RBF	Acc: $92.1 \pm 1.8\%$	$90.1 \pm 2.6\%$
	$C = 55$	Sen: $88.3 \pm 1.2\%$	$84.5 \pm 5.1\%$
	$\gamma = 7$	Spe: $94.1 \pm 2.1\%$	$93.2 \pm 3.3\%$

Probabilistic cancer colormaps: Prior to the generation of each colormap, the SVM was trained on all the training data except for the ROIs originating from the patient under study. Figure 1 demonstrates two of the colormaps acquired using the described seven-dimensional feature vector and the SVM-based probabilistic outcomes. The colormaps were generated by tissue typing of the entire peripheral zone of the prostate of patients. The colored ROIs in the colormaps were those with $P_c > 0.4$ (See Section 2.3). The choice of the threshold was rather subjective, however, we found that with the threshold of $P_c > 0.4$, the visualized maps demonstrated all the major tumors in the dataset without a significant number of false positive detections.

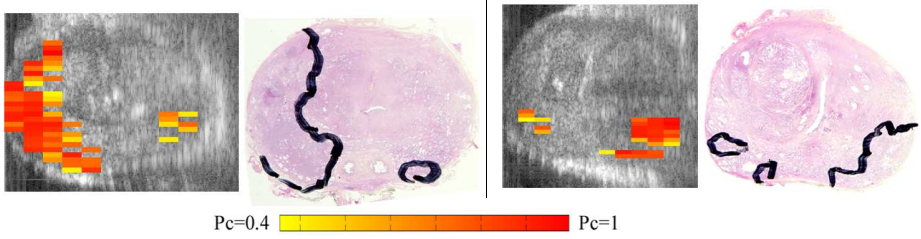


Fig. 1. Cancer probability maps along with the closest matching histopathology slide. The probe is in touch with the left side of the gland.

Quantitative validation: Colormaps can be viewed as qualitative measures of the performance of the technique. To quantitatively validate the authenticity of the technic, we performed patient-based classification experiments. In these experiments, the test data consisted only of the ROIs for which the pathological condition was confidently known. For each patient, we trained the SVM based on ROIs from all other patients and tested it on ROIs from the patient in question. With threshold probability of cancer set to $P_c = 0.4$, an average accuracy of 79% with sensitivity of 81% and specificity of 77% were achieved (area under ROC curve = 0.85). We noticed that when only texture and LF features were used, the patient-based accuracy dropped to 75% and more importantly, the sensitivity was only 68% (area under ROC curve: 0.79). In other words, the time series features are significant contributors to sensitivity of the colormaps.

4 Discussions

The effect of tissue depth: Our analysis showed that tissue typing based on RF time series features is more accurate in areas that are closer to the ultrasound probe. We witnessed a meaningful drop in the accuracy of our tissue classification (based on RF time series features) for ROIs that are more than 3 cm away from the probe (Figure 2). With the only focal point set at 2 cm, the ultrasound beams diverge increasingly at deeper areas of the tissue. Furthermore, the effects of amplifier noise are more prevalent in deeper areas. Therefore, data from the deeper areas suffer from an increasing level of noise¹. The tissue typing information is carried in time series variations, and the increasing levels of destructive noise in deeper areas of the tissue decreases the signal to noise ratio (SNR) in the received RF time series. From the clinical point of view, the depth-dependent accuracy of the approach does not pose a significant challenge. The reason is that the most cancer-prone area of the prostate gland is the peripheral zone which is also the target of inspection during biopsy.

The effect of ultrasound power: A reduction in the acoustic transmit power of the ultrasound probe also reduces the SNR in the RF time series. The ultrasound

¹ Noise is measured as $\sum_{i=1}^{i=112} (x_i - \bar{x})^2$ where \bar{x} is the mean of the RF time series and x_i 's are the samples of the RF time series; the length of the time series is 112.

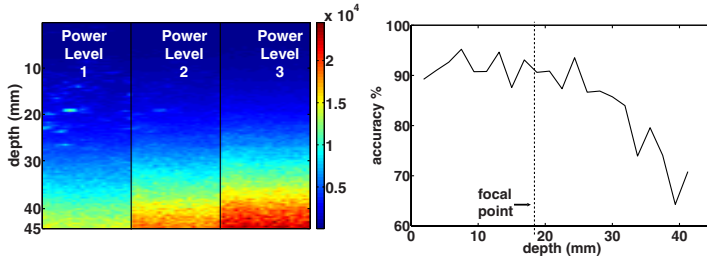


Fig. 2. Right: The accuracy of tissue typing based on time series *vs.* depth. Left: Noise at different depths for three transducer power settings ($0dB$, $-2dB$ and $-6dB$).

machine is normally operated at its highest acoustic power level to maximize the contrast and SNR. A $2dB$ reduction in the output power resulted in a drop of accuracy from $82.5 \pm 1.6\%$ to $80.4 \pm 3.1\%$. A $6dB$ power reduction caused a drop in the accuracy of the method to $70.1 \pm 3.9\%$. Figure 2 demonstrates a visual comparison of the noise levels at the three studied power settings.

5 Conclusions

The results show that the RF time series features are powerful tissue typing parameters. They are significantly more accurate and sensitive than RF spectral features [3] and also B-scan texture features. These two approaches complement the RF time series method resulting in a very effective feature vector that provides an area under ROC curve of 0.95 in 10-fold cross-validation. Based on this hybrid feature vector and an extension to the SVM classification approach [12], we created colormaps that accurately highlighted areas of tissue with high risk of cancer on ultrasound images.

The analysis of RF time series is computationally expensive compared to the analysis of single RF or B-scan frames. However, our implementations show that the feature values can be extracted from a 1 cm^2 area, in less than a second on a single CPU computer. The proposed method involves very limited overhead to the routine clinical examination. Nevertheless, the radiologist needs to keep a hand steady for about 3-5 seconds while the RF time series are acquired. We are planning *in-vivo* studies that will determine the effectiveness of the method in clinical practice and analyze the effects of potential hand movements. Also, we are currently performing phantom studies that will measure the sensitivity of RF time series features to changes in elasticity and scatterer size.

References

1. Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Thun, M.: Cancer statistics, 2007. CA: A Cancer Journal for Clinicians (57), 43–66 (2007)
2. Moradi, M., Mousavi, P., Abolmaesumi, P.: Computer-aided diagnosis of prostate cancer with emphasis on ultrasound-based approaches: A review. Ultrasound in Medicine and Biology 33(7), 1010–1028 (2007)

3. 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 spectrum analysis. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 43(4), 609–619 (1996)
4. Scheipers, U., Ermert, H., Garcia-Schurmann, H.J.S.M., Senge, T., Philippou, S.: Ultrasonic multifeature tissue characterization for prostate diagnosis. *Ultrasound in Medicine and Biology* 20(8), 1137–1149 (2003)
5. Ophir, J., Cespedes, I., Ponnekanti, H., Yazdi, Y., Li, X.: Elastography: a method for imaging the elasticity in biological tissues. *Ultrasonic Imaging* 13(2), 111–134 (1991)
6. Moradi, M., Mousavi, P., Siemens, D.R., Sauerbrei, E.E., Isotalo, P., Boag, A., Abolmaesumi, P.: Discrete Fourier analysis of ultrasound RF time series for detection of prostate cancer. In: *IEEE EMBC*, pp. 1339–1342 (2007)
7. Moradi, M., Mousavi, P., Abolmaesumi, P.: Tissue characterization using fractal dimension of high frequency ultrasound RF time series. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II. LNCS*, vol. 4792, pp. 900–908. Springer, Heidelberg (2007)
8. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
9. Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena* 31(2), 277–283 (1988)
10. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer Science, New York (2006)
11. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using the second order information for training SVM. *Machine Learning Research* 6, 1889–1918 (2005)
12. Platt, J.C.: Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: *Advances in Large Margin Classifier*. MIT Press, Cambridge (2000)