Prostate Cancer: Improved Tissue Characterization by Temporal Modeling of Radio-Frequency Ultrasound Echo Data

Layan Nahlawi[1], Farhad Imani[2], Mena Gaed[3], Jose A. Gomez[3], Madeleine Moussa[3], Eli Gibson[4], Aaron Fenster[3], Aaron D. Ward[3], Purang Abolmaesumi[2], Hagit Shatkay[1,5⋆], and Parvin Mousavi[1⋆]

[1] School of Computing, Queen's University
[2] Dept. of Electrical and Computer Engineering, University of British Columbia
[3] Centre for Medical Image Computing, University College London
[4] Dept. of Medical Biophysics, Pathology and Robarts Inst., Western University
[5] Dept. of Computer and Information Sciences, University of Delaware

Abstract. Despite recent advances in clinical oncology, prostate cancer remains a major health concern in men, where current detection techniques still lead to both over- and under-diagnosis. More accurate prediction and detection of prostate cancer can improve disease management and treatment outcome. Temporal ultrasound is a promising imaging approach that can help identify tissue-specific patterns in time-series of ultrasound data and, in turn, differentiate between benign and malignant tissues. We propose a probabilistic-temporal framework, based on hidden Markov models, for modeling ultrasound time-series data obtained from prostate cancer patients. Our results show improved prediction of malignancy compared to previously reported results, where we identify cancerous regions with over 88% accuracy. As our models directly represent temporal aspects of the data, we expect our method to be applicable to other types of cancer in which temporal-ultrasound can be captured.

1 Introduction

Prostate cancer is the most widely diagnosed form of cancer in men [1]. The American Cancer Society predicts that one in seven men will be diagnosed with prostate cancer during their lifetime. Initial assessment includes measuring Prostate Specific Antigen level in blood serum and digital rectal examination. If either test is abnormal, core needle biopsy is performed under Trans-Rectal Ultrasound (TRUS) guidance. Disease prognosis and treatment decisions are then based on grading, i.e., assessing the degree of cancer-aggressiveness in the biopsy cores. TRUS-guided biopsy often leads to a high rate (∼30%) of false negatives for cancer diagnosis as well as to over- and under-estimation of the cancer grade. Extensive heterogeneity in morphology and pathology of prostate adenocarcinoma are additional challenging factors for making an accurate diagnosis.

While improved prostate-cancer screening has reduced mortality rates by

⋆ These authors have contributed equally to the manuscript.

45% over the past two decades [3], inaccurate diagnosis and grading have resulted in a surge in over-treatment. Notably, radical over-aggressive treatment of prostate-cancer patients leads to a decline in their quality of life. For indolent prostate cancer, such aggressive treatment should be avoided as active surveillance has proven to be an effective disease management course [13]. Accurate identification and grading of lesions and their extent – especially using low cost, readily accessible technology such as ultrasound – can, therefore, significantly contribute to appropriate effective treatment. To achieve this, methods must be developed to guide TRUS biopsies to target regions likely to be malignant. The task of differentiating malignant tissue from its surrounding tissue is referred to in the literature as *tissue-typing* or *characterization*. In this paper we propose a new method that utilizes ultrasound time-series data to characterize malignant vs. benign tissue obtained from prostate-cancer patients.

Most of the research on ultrasound-based tissue characterization focuses on analysis of texture- [5] and spectral-features [4] within single ultrasound frames. Elastography [8], another ultrasound technique, aims to distinguish tissue types based on their measured stiffness in response to external vibrations. A different way of utilizing ultrasound is by acquiring radio-frequency (RF) time series, which is a sequence of ultrasound frames captured from sonication of tissue over time, without moving the tissue or the ultrasound probe. Frequency domain analysis of RF time series has shown promising results for tissue characterization in breast and prostate cancer. Moradi et al. [10] used the fractal dimension of RF time series as features and employed Bayesian and neural network classifiers for ex-vivo characterization of prostate tissue. More recently, Imani et al. [7] combined wavelet features and mean central frequency of RF time-series to characterize in-vivo prostate tissue using SVMs. Neither of these lines of work have explicitly modeled the temporal aspect of the time-series data and utilized it for tissue characterization. In a recent study [11] we have suggested that the temporal aspect of the data may carry useful information if directly captured.

Here we carry the idea forward, presenting a new method for analyzing RF time series, using a probabilistic temporal model, namely, a hidden Markov model, HMM [12], specifically representing and making use of the temporal aspect of the data. We apply the method to differentiate between malignant and benign prostate tissue and demonstrate its utility, showing an improved performance compared to previous methods. Probabilistic temporal modeling, and HMMs in particular, have been applied to a wide range of clinical data. They are typically used to model a time-dependent physiological process (e.g. heartbeats [2]), or the progression of disease-risk over time [6]. HMMs are also used within and outside the biomedical domain to model sequence-data such as text [9], proteins, DNA sequences and others. Here we use them to model RF time series where time does have an impact on the ultrasound data being recorded. We next describe our RF time-series data and its representation, followed by a tissue-characterization framework. We then present experiments and results demonstrating the effectiveness of the method.

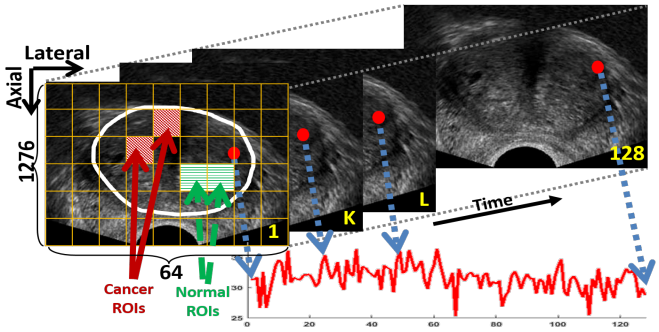

Fig. 1: Ultrasound RF-frames collected from a prostate-cancer patient over time. Solid red dots indicate the same location across multiple frames. The time series for this location is shown at the bottom right. A grid dividing each frame into ROIs is shown on the left-most frame. Pathology labels for malignant/benign ROIs are also shown.

2 RF Time Series Data

RF time series record tissue-response to prolonged sonication. These responses consist of reflected ultrasound echo intensity values. Fig. 1 shows ultrasound image-frames collected from prostate sonication over time (each such frame is referred to as an RF *frame*). The boundary of the prostate is encircled in white. The solid red dots indicate the same location within the prostate over time, while the dotted blue arrows point to the corresponding echo intensity values. The sequence of echo intensities obtained from the same point in the prostate over time makes up an RF time series (bottom right of the Figure). Due to the scattering phenomenon in ultrasound imaging, very small objects such as individual cells cannot be identified using single RF values. As such, we partition each RF frame using a grid into small regions, where each window in the grid is referred to as a *Region of Interest (*ROI*)*, and comprises multiple RF values. In this work, we use the same dataset as Imani et al [7], and adopt the same ROI size 1.7x1.7mm^2, which corresponds to 44x2 RF values. The 88 RF values within each grid-window in a single frame recorded at time t, are averaged to produce a single value representing each ROI at the corresponding time-point t.

The image data consists of in-vivo RF frames gathered from 9 prostate-cancer patients who have undergone radical prostatectomy[6]. Prior to the surgery, 128 RF frames recorded over a time-period of 4 seconds, were gathered from each patient. A grid is overlaid on each of the frames, ROIs are obtained as described above, and for each ROI, $\mathcal{R}_k$, a 128-long time series $\mathcal{R}_k = \langle {R^k}_1, \ldots, {R^k}_{128} \rangle$ is created. Each point ${R^k}_t$ in the series corresponds to the average intensity of that ROI in the RF-frame recorded at time t, where $1 \leq t \leq 128$. While the number of patients is relatively low, the total number of ROIs per patient is quite high (see Table 1), thus providing sufficient data to support effective model-learning.

As commonly done in time series analysis, we map the series associated with each ROI, $\mathcal{R}_k$, to its first-order difference series, i.e. the sequence of differences

[6] The study was approved by the institutional research ethics board and the patients provided consent to participate.

Table 1: The distribution of malignant and benign ROIs over the 9 patients.

Patient	P1	P2	P3	P4	P5	P6	P7	P8	P9	Total
Malignant ROIs	42	29	18	64	35	28	23	30	17	286
Benign ROIs	42	29	18	61	35	29	23	30	17	284

between pairs of consecutive time-points. To simplify the modeling task, we further discretize the difference series, by placing the values into 10 equally-spaced bins, where the values in the lowest bin are all mapped to 1, and those at the top-most bin are mapped to 10. We denote the sequence obtained by discretizing $\mathcal{R}_k, as\ \langle {O^k}_1, ..., {O^k}_{128} \rangle$. Our experiments suggest that 10 bins are sufficient for obtaining good classification performance.

To create a gold-standard of labeled *malignant* vs *benign* regions we use whole-mount histopathology information. To obtain such information, following prostatectomy, the tissues are formalin-fixed and imaged using MRI. The tissues are then cut into ∼4mm slices, and further processed to enable high resolution microscopy. Two clinicians then assign (in consensus) the appropriate labels to the malignant and to the benign areas within each slice. A multi-step rigorous registration process, in which MRI images are used as an intermediate step, is employed to overlay the labeled histopathology images on the in-vivo ultrasound frames (see [7] for additional details). This registration process results in an assignment of a pathology label to each ROI, indicating whether it is malignant or benign. Fig. 1 shows several examples of such labeled ROIs. We use the same 570 labeled ROIs as in [7], of which 286 are malignant and 284 benign. Table 1 summarizes the data. The RF time-series associated with the labeled ROIs are used as training and test data for building a probabilistic model for distinguishing between benign and malignant tissues, as described in the next section.

3 Probabilistic Modeling Using Hidden Markov Models

HMMs are often used to model time series where the generating process is unknown or prone to variation and noise. The process is viewed as a sequence of stochastic transitions between unobservable (hidden) states; some aspects of each state are observed and recorded. As such, the states may be estimated from the observation-sequence [12]. A simplifying assumption underlying the use of these models is the Markov property, namely, that the state at a given time-point depends only on the state at the preceding point, conditionally independent of all other time points. In this work we view a tissue response value recorded in an RF frame and discretized as discussed above, as an observation; employing the Markov property, we assume each such value depends only on the response recorded at the frame directly preceding it, independent of any earlier responses. Formally, an HMM λ consists of five components: A set of N states, $S = \{s_1, \ldots, s_N\}$; a set of M observation symbols, $V = \{v_1, \ldots, v_M\}$; an NxN stochastic matrix A governing the state-transition probability, where $A_{ij} = Pr(state_{t+1} = s_i | state_t = s_j),\ 1 \leq i, j \leq N$, and $state_t$ is the state at time t; an NxM stochastic-emission matrix B, where $B_{ik} = Pr(ob_t = v_k | state_t = s_i)$, $1 \leq i \leq N, 1 \leq k \leq M$, denoting the probability of observing v_k at state s_i; an

N-dimensional stochastic vector π, where for each state s_i, $\pi_i = Pr(state_1 = s_i)$, denotes the probability to start the process at state s_i. *Learning a model* λ from a sequence of observations $O = o_1, o_2, \ldots, o_{128}$, amounts to estimating the model parameters (namely, A, B & π), to maximize $\log[Pr(O|\lambda)]$, i.e. the observations' probability given the model λ. In practice, π is fixed such that $\pi_1 = Pr(state_1 = s_1) = 1$ & $\pi_j = 0$ for $j \neq 1$, i.e. s_1 is always the first state. In the experiments reported here, we also fix the matrix A to an initial estimate based on clustering (as described below), while the matrix B is learned using the Baum-Welch algorithm [12].

The HMMs we develop, as illustrated in Fig. 2, are ergodic models consisting of 5 states and 10 observations. A small number of states allows for a computationally efficient model while typically leading to good generalization beyond the training set. We determined the number of states by experimenting with 2-6 state models (and a few larger ones with > 10 states). The classification performance of 5-state models was higher than that of others. Moreover, each of the 5 states is associated with a distinct emission probability distribution, which is not the case when using additional/fewer states. The observation set, as discussed in section 2, consists of 10 observation symbols $v_1, ..., v_{10}$, each of which corresponds to a discretized interval of first-order difference values of the RF time-series.

For tissue classification, we learn two HMMs – one for representing series obtained from *malignant* tissue, denoted λ_M, and the other for *benign* tissue, denoted λ_B. We use supervised learning to learn the models' parameters, where the training and test data consist of the time-series corresponding to the ROIs that were labeled as *malignant* and *benign* (described in Sec. 2). To train each model, we use a leave-one-patient-out cross-validation strategy, partitioning each set of ROI time-series (*malignant* for λ_M, *benign* for λ_B) into training and test sets. In each cross-validation run the ROIs of one of the 9 patients are left-out as a test-set, while the ROIs of the other 8 patients are used to train the HMM. Malignant ROIs are used to train λ_M, while λ_B is trained on benign ROIs. Given a test-sequence, roi_{test_i}, each of the two models assigns it a log probability, $log(Pr(ROI_{test_i}|\lambda_c))$, $(c \in \{M, B\})$ – a measure indicating how likely the model is to have generated the time-series. The class label assigned to ROI_{test_i}, C_{test_i}, is the one whose model maximizes the log probability, that is:

$C_{test_i} = \arg\max_{c \in \{M,B\}}(log(Pr(ROI_{test_i}|\lambda_c)))$, $1 \leq i \leq L$, where L is the # of test-ROIs. Practically, if the log-odds $log\frac{Pr(ROI_{test_i}|\lambda_B)}{Pr(ROI_{test_i}|\lambda_M)}$ is positive, ROI_{test_i} is classified as *malignant*, otherwise it is classified as *benign*. In Section 4, we use the log-odds as a basis for heat-maps to visualize the results (Fig. 3).

To learn the two models, each of the models is initialized, and its observation matrix B is then iteratively updated until convergence, in accordance with the Baum-Welch method. Model initialization is based on clustering the values within all the discretized training time-series into 5 clusters, $cl_1, \ldots, cl_5$, where 5 is the number of states. Based on the assignment of each value to its respective cluster, we estimate the transition probability $A_{i,j}$ where $1 \leq i, j \leq 5$ as the

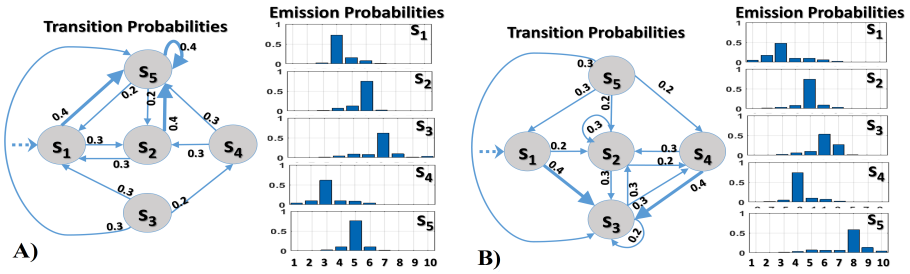

Fig. 2: Example of HMMs learned from A) malignant ROIs, and B) benign ROIs. Nodes represent states. Edges are labeled by transition probabilities; Emission probabilities are shown to the right of each model. Edges with probability ≤ 0.2 are not shown.

data-frequency of observing a value from cluster cl_i followed by a value from cluster cl_j within all the time series in the respective training set. Since the model is not left-to-right, the transitions can be in either direction. A similar estimation process is applied for initializing the observation matrix B.

To assess our method's performance, we apply each of the trained models (trained over ROI time-series obtained from 8 patients) to assign labels over the test data (the ROIs of the left-out patient), and calculate the average standard measures *accuracy, sensitivity* and *specificity* of the assigned labels with respect to the ground-truth in the gold-standard. The learnt HMMs provides a summary of the course of changes in RF values that each of the tissue types goes through in response to the prolonged sonication.

4 Results and Discussion

As explained above, we train 9 pairs of HMMs – one malignant and one benign – where for each pair, the training is done on data obtained from 8 of the 9 patients. Each pair of HMMs is then used for classifying the ROIs of the left-out 9^{th} (test) patient. Fig. 2 shows an example of such a pair of learned HMMs, where the left one was trained on time-series obtained from malignant ROIs while the one on the right was trained on benign ones. The transition probabilities are shown on the edges while emission probabilities for each state are shown as histograms. The figure shows that in both models, each state is characterized by its own markedly distinct observation distribution. Moreover, the most likely

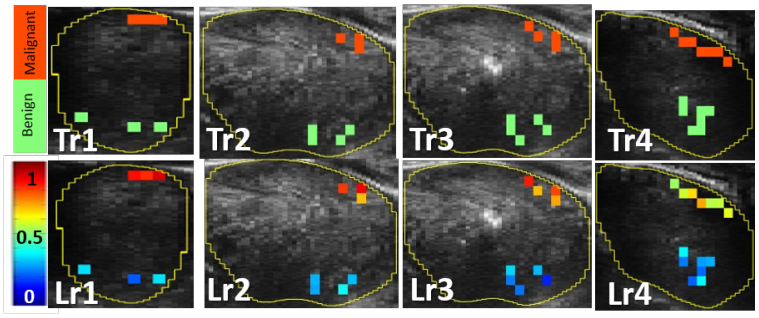

Fig. 3: Top: RF frames overlaid with malignant/benign pathology labels. Bottom: Heat-map images based on our learned models, where each ROI color is assigned based on the log-odds ratio calculated for its respective time-series. The left three columns are RF frames from patients *P1* (col 1) and *P5* (col 2, 3) while the frames in the rightmost column are from Patient *P7*, for whom we noted a lower performance.

Table 2: The classification performance using HMMs. The numbers in parentheses show the respective result reported by [7] for the same patient.

Patient	P1	P2	P3	P4	P5	P6	P7	P8	P9	Average
Accuracy	82.1 (82)	96.5 (71)	100 (88)	93.6 (95)	90 (86)	85.9 (86)	69.5 (N/A)	78.3 (80)	97.1 (85)	88.1 ±9
Sensitivity	100 (100)	96.5 (68)	100 (76)	87.5 (90)	97.1 (100)	82.1 (81)	65.2 (N/A)	73.3 (98)	100 (84)	89.1 ±12
Specificity	64.2 (62)	96.5 (74)	100 (100)	100 (100)	82.8 (71)	89.6 (90)	73.9 (N/A)	83.3 (61)	94.1 (84)	87.1 ±11

path transitioning through the malignant model alternates primarily between the states s_1 and s_5 possibly via s_2, where s_5 is the most central state, that is, the one most likely to be visited. In contrast, the benign model alternates primarily between the states s_1, s_3 and s_4, with s_3 being the most central state. Notably the emission distribution associated with s_5 in the malignant model is very different from that associated with s_3 in the benign one, hence these two states are not equivalent. The clear distinction between the two models means that time series obtained from malignant ROI's form a certain typical pattern of changing values, while time series obtained from benign ROI's form a different typical pattern, and our models do capture the difference.

The classification results for all test patients are shown in Table 2. The average accuracy is 88.1%, whereas the average sensitivity and specificity are 89.1% and 87.1%, respectively. The results clearly indicate that for the majority of ROIs, our trained models can correctly distinguish between ROIs obtained from malignant tissue and those obtained from benign tissue. Moreover, for most cases our performance either matches or significantly improves upon that of an earlier method [7] that used SVMs and did not explicitly model the temporal aspect of the time-series. We note that for patient *P8* our sensitivity is significantly lower, although our specificity is much higher, which amounts to a comparable overall accuracy. An exception to the high level of performance is clearly observed for patient *P7*, for whom the classification performance is significantly lower than that obtained for all other patients. Further investigation showed that this patient was not included in the earlier reported results [7], because the ground-truth registration of the histology labels of malignant tissue was not accurate. The fact that mis-labeled ROIs are not well-distinguished based on the models learned from other patient data serves as further evidence for the fact that the models indeed capture the salient differences between RF echos emitted by benign vs malignant tissue. The top row of Fig. 3 shows several examples of RF frames obtained from different patients overlaid with malignant/benign labels. The bottom row shows corresponding images of heat-maps based on our results. Each ROI, is assigned a color reflecting the log-odds ratio calculated for its respective time-series $\mathcal{R}_x$, $log\frac{Pr(\mathcal{R}_x|\lambda_B)}{Pr(\mathcal{R}_x|\lambda_M)}$. The first three columns show RF frames from *P1* (1^{st} column) and *P5* (2^{nd}, 3^{rd} columns), all of which show that the heat-maps match the original annotations almost perfectly. The fourth column shows an RF frame from *P7*. Despite inaccuracies in the gold-standard for this image, our model still identifies correctly the benign regions, while showing most of the malignant regions about equally likely to be malignant or benign.

5 Conclusion

We introduced a new approach for tissue-classification in prostate cancer, based on modeling temporal aspects of tissue-response to prolonged sonication. Representing the two tissue types (*malignant*/*benign*), each as a probabilistic-temporal model learned from patients' data (training-set) allows for accurate labeling of test-data obtained from another patient. Our results indicate that temporal patterns, captured by our models, help differentiate between RF time series obtained from malignant vs. benign tissues, with an average accuracy of over 88%. As a next step we plan to take into account the heterogeneity in benign tissue, as well as incorporate cancer grades to support a more refined categorization of tissue types. This study takes a first step using such models, and is limited by a relatively small number of patients for which we have reliably annotated whole-mount tissue images. In the future we shall increase the number of patients, and include anatomical-data indicating the zones from which ROIs are selected. Beyond prostate cancer, we expect our method to be applicable to other types of cancer such as breast and liver.

Acknowledgment

This work was partially supported by grants from NSERC Discovery to HS and PM, NSERC and CIHR CHRP to PM and NIH #R56 LM011354A to HS.

References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics: Canadian Cancer Statistics 2015. Toronto, ON. Can Cancer Soc (2015). ISSN 0835-2976
2. Coast, D., Stern, R., et al.: An approach to cardiac arrhythmia analysis using HMM. IEEE Trans on Biomed Eng 37(9), 26–36 (1990)
3. Etzioni, R., Tsodikov, A., et al.: Quantifying the role of psa screening in the us prostate cancer mortality decline. Cancer Causes & Control 19(2), 75–81 (2008)
4. Feleppa, E., Porter, et al.: Recent developments in tissue-type imaging for planning and monitoring treatment of prostate cancer. Ultrasonic Img 26(3), 63–72 (2004)
5. Han, S., Lee, H., et al.: Computer-aided prostate cancer detection using texture features and clinical features in US image. J Dig Img 21(1), 21–33 (2008)
6. Hauskrecht, M and Fraser, H.: Planning treatment of ischemic heart disease with partially observable Markov decision processes. AI in Medicine 18(3), 21–44 (2000)
7. Imani, F., Abolmaesumi, P., et al.: Computer-aided prostate cancer detection using US RF time series: In vivo feasibility study. IEEE Trans on Med Img 34(11), 48–57 (2015)
8. Krouskop, T., Wheeler, T., et al.: Elastic moduli of breast and prostate tissues under compression. Ultrasonic Img 20(4), 60–74 (1998)
9. Li, Y., Lipsky Gorman, S., et al.: Section classification in clinical notes using supervised HMM. Proc of the ACM Int. Health Informatics Symp. 44–50 (2010)
10. Moradi, M., Mousavi, P., et al.: Augmenting detection of prostate cancer in TRUS images using SVM and RF time series. IEEE Trans on Biomed Eng 56(9), 214–224 (2009)
11. Nahlawi, L., Imani, F., et al.: Using hidden markov models to capture temporal aspects of US data in prostate cancer. IEEE Int Conf on BIBM 2015. 46–49 (2015)
12. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proc of the IEEE 77(2), 257–286 (1989)
13. Singer, E.A., Kaushal, A., et al.: Active surveillance for prostate cancer: past, present and future. Current Op in Oncology 24(3), 43–50 (2012)