Ultrasound-Based Detection of Prostate Cancer Using Automatic Feature Selection with Deep Belief Networks

Shekoofeh Azizi[1], Farhad Imani[1], Bo Zhuang[1], Amir Tahmasebi[2], Jin Tae Kwak[3], Sheng Xu[3], Nishant Uniyal[1], Baris Turkbey[4], Peter Choyke[4], Peter Pinto[4], Bradford Wood[4], Mehdi Moradi[5], Parvin Mousavi[6], and Purang Abolmaesumi[1]

[1] The University of British Columbia, Vancouver, BC, Canada
[2] Philips Research North America, Briarcliff Manor, NY, USA
[3] National Institutes of Health, Bethesda, MD, USA
[4] National Cancer Institute, Bethesda, MD, USA
[5] IBM Almaden Research Center, San Jose, CA, USA
[6] Queen's University, Kingston, ON, Canada

Abstract. We propose an automatic feature selection framework for analyzing temporal ultrasound signals of prostate tissue. The framework consists of: 1) an unsupervised feature reduction step that uses Deep Belief Network (DBN) on spectral components of the temporal ultrasound data; 2) a supervised fine-tuning step that uses the histopathology of the tissue samples to further optimize the DBN; 3) a Support Vector Machine (SVM) classifier that uses the activation of the DBN as input and outputs a likelihood for the cancer. In leave-one-core-out cross-validation experiments using 35 biopsy cores, an area under the curve of 0.91 is obtained for cancer prediction. Subsequently, an independent group of 36 biopsy cores was used for validation of the model. The results show that the framework can predict 22 out of 23 benign, and all of cancerous cores correctly. We conclude that temporal analysis of ultrasound data can potentially complement multi-parametric Magnetic Resonance Imaging (mp-MRI) by improving the differentiation of benign and cancerous prostate tissue.

Keywords: Temporal ultrasound data, deep learning, deep belief network, cancer diagnosis, prostate cancer, feature selection, classification.

1 Introduction

The early diagnosis of prostate cancer, as the most common type of diagnosed malignancy in North American men, plays an important role in the choice and the success of treatments [8]. The definitive diagnosis of prostate cancer is histopathological analysis of biopsy tissue samples, which is typically guided under Transrectal Ultrasound (TRUS). However, conventional systematic biopsy under TRUS guidance does not have high sensitivity and specificity. Recently,

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 70–77, 2015.
DOI: 10.1007/978-3-319-24571-3_9

fusion of multi-parametric MRI (mp-MRI) with TRUS has shown significant potential for improved cancer yield [7]. A meta-analysis of seven mp-MRI studies with 526 patients show specificity of 0.88 and sensitivity of 0.74, with negative predictive values ranging from 0.65 to 0.94 [9]. mp-MRI suffers from several limitations including low sensitivity for detection of small lesions and low grade cancer. Hence, there is a need to develop new imaging techniques that can complement mp-MRI for prostate cancer diagnosis.

Over the past two decades, several ultrasound-based techniques have been proposed for characterizing cancerous tissue [8]. The clinical uptake of these methods has been slow, mainly due to the large variability of the tissue characterization results. The primary sources of such variability are the heterogeneous patient population and cancer grades. Moreover, these methods do not generally report clinically sufficient specificity and sensitivity to ensure that all aggressive cancer cases can be captured by ultrasound analysis.

Ultrasound-based tissue typing techniques to-date mainly rely on spectral analysis of a single ultrasound image, or alternative modalities such as elastography and Doppler, that are not commonly used during prostate biopsy. Recently, analysis of temporal ultrasound data obtained from a fixed tissue location has been proposed to complement other ultrasound-based tissue typing techniques [4]. The key advantage of this technique is that it relies on conventional ultrasound imaging data, without the need to change the clinical scanning protocol. It has been shown that temporal changes in the ultrasound signal backscattered from the tissue varies across tissue types. Several hypotheses have been explored to determine the physical phenomenon governing the approach. It has been proposed that the acoustic radiation force of the transmit ultrasound signal increases the tissue temperature, and changes the speed of sound [2]. Another hypothesis indicates that micro-vibration of acoustic scatterers, contribute to tissue typing in this approach. Previous works [4,5,12] utilize a machine learning framework to determine the correlation between features extracted from the temporal ultrasound data and cancer presence or grade provided by histopathology. In these methods, features were heuristically determined from the spectral/wavelet analysis of the ultrasound data. In addition to challenges associated with defining features that are correlated with the underlying tissue properties, it is also difficult to determine the best combination of those features for effective tissue typing as the number of features increases. The lack of a systematic approach for feature selection can lead to a so-called "cherry picking" of the features [4,5,12].

Deep learning approaches have gained significant attention for capturing abstract representations of input data [1], and have been successfully used in medical image analysis [6,11]. In this paper, we exploit Deep Belief Networks (DBN) [1] to automatically learn a high-level feature representation of the temporal ultrasound data that can detect prostate cancer. Subsequently, we use the high-level features in a supervised classification step based on Support Vector Machines (SVM) to generate a cancer likelihood. In a study consisting of 71 biopsy cores obtained from 62 subjects, we demonstrate that the approach is

effective in identifying both benign and cancerous biopsy cores in TRUS-guided biopsy, and can complement mp-MRI to further improve the cancer yield.

2 Materials

Temporal ultrasound Radio-Frequency (RF) data was acquired with a Philips iU22 ultrasound scanner during MRI-guided targeted TRUS biopsies using the Philips UroNav platform at the National Institutes of Health Clinical Center (NIH-CC), Bethesda, MD. Prior to the biopsy session, mp-MR images including T2-weighted, Diffusion Weighted Images (DWI), and Dynamic Contrast Enhanced (DCE) images were acquired for each subject. Two independent radiologists identified and scored suspicious lesions in MR images based on a standard protocol, ranging from 1 (no cancer) to 5 (aggressive cancer) and assigned grades: low (2 or less), moderate (3) and high (4 or more). The desired biopsy targets were delineated on the T2-weighted MR image based on these scores. In the beginning of the procedure, a 3D ultrasound volume of the prostate was reconstructed by acquiring a series of electromagnetically (EM) tracked 2D TRUS images. For targeted biopsy, T2-weighted MR images were automatically fused with the 3D US volume of the prostate using the UroNav platform. Following the registration of ultrasound and MR volumes, the targeted locations for biopsy were transformed to the EM coordinate frame. During biopsy, the clinician navigates through the prostate volume to reach the desired target location for acquiring a core. The clinician then holds the TRUS transducer fixed for about 5 seconds to acquire 100 frames of ultrasound time series data. Temporal ultrasound data is collected from at most two MR-identified targets per subject. Tissue biopsy is then followed by histopathology analysis and results are used as the ground-truth labels for evaluation of cancer detection. The level of abnormality for each biopsy core is described using Gleason Score (GS). Data is obtained from 71 biopsy cores of 62 subjects; among them 45 cores are benign while 26 other cores are cancerous. The data was divided into two independent data sets for training, D_1, and validation, D_2. For building the model, we need to use homogeneous tissue regions. Given potential mis-registration errors between MR and ultrasound images, we only include biopsy cores with at least 7.2 mm of cancer for a typical core length of 18 mm. We also add benign cases to the training data set.

The training dataset, D_1, consists of 35 biopsy cores obtained from 31 patients with the following histopathology label distribution: Benign: 22; GS 3+3: 0; GS 3+4: 5; GS 4+3: 2; GS 4+4: 4; and, GS 4+5: 2 cores. The validation data set, D_2, contains another 36 biopsy cores obtained from 31 patients with the following histopathology label distribution: Benign: 23; GS 3+3: 4; GS 3+4: 3; GS 4+3: 0; GS 4+4: 4; and, GS 4+5: 2 cores.

We extract features from temporal ultrasound data over 20 equally-sized Regions of Interest (ROI) of size 1 mm^2 for each biopsy core. For each biopsy target, we analyze an area of 2×10 mm^2 in the lateral and axial directions, respectively, along the projected needle path in the ultrasound image, and centered on the

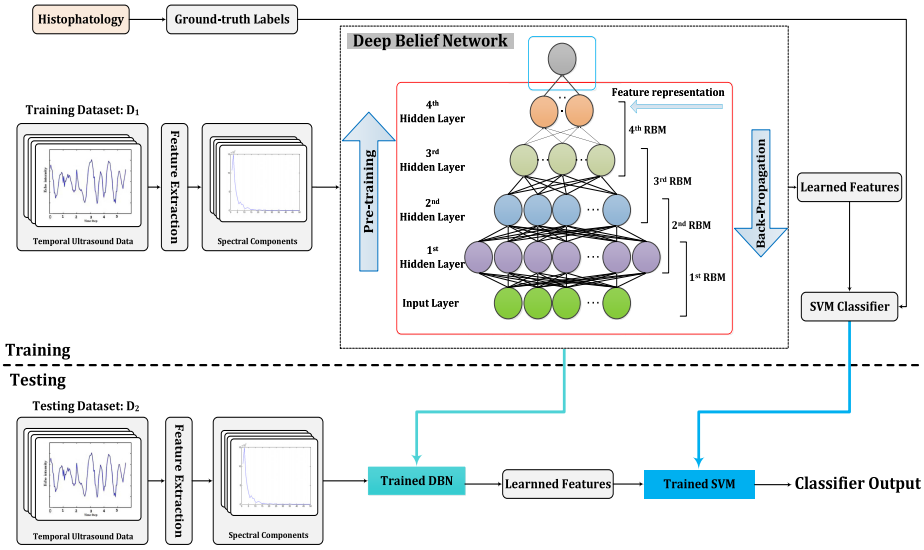

Fig. 1. An illustration of the proposed method for prostate cancer detection. Our DBN has a layer of real-valued visible units of dimension 50 and four hidden layers with 100, 50, 45 and 6 hidden units. The red box contains the pre-trained DBN, and the blue box containing one neuron is added for the fine-tuning step. The latent features are the output of the last layer of DBN.

target. The width of this area is close to the width of the biopsy core, but the length of the biopsy core is typically 18 mm. To generate the features of temporal ultrasound data, we take the Fourier transform of the entire time series, normalized to the frame rate. For each ROI, we generate 50 spectral features by averaging the absolute values of the Discrete Fourier Transform (DFT) of the zero-mean ultrasound time series for 50 positive frequency components. Due to symmetry, only the positive frequency components of the spectrum are used.

3 Methods

Fig. 1 illustrates a schematic diagram of the proposed method. Using D_1, we derive high level - latent- feature representation from the 50 extracted features using DBN. Then, we use the learned feature information along with ground-truth labels to train a Support Vector Machine (SVM). To verify the generated model, we extract the features for data set D_2. Then, we use the activations of the trained DBN and SVM to derive the tissue type labels for each ROI.

3.1 Deep Belief Networks

DBN, as a generative probabilistic model, can be trained to extract a deep hierarchical representation of the input data. We construct a DBN by stacking

many layers of Restricted Boltzmann Machines (RBMs), each of which consists of a visible layer and a hidden layer of neurons. The two layers are connected by a matrix of weighted connections and there are no connections within a layer. RBMs can be stacked by linking the hidden layer of one RBM to the visible layer of the next RBM. A schematic representation is shown in Fig. 1. The role of a RBM is to model the distribution of its input and capturing patterns in the visible units. Thus, DBNs offer an unsupervised way to learn multi-layer probabilistic representations of data that are progressively *deeper* with each successive layer. Utilizing this representational power, we can find a latent representation of the original low-level features extracted from temporal ultrasound data.

We perform the contrast divergence approximation [1] to update the weights during training. The DBN can be trained by performing a greedy layer-wise unsupervised *pre-training* phase [1]. In a greedy layer-wise learning, we train one layer at a time and use the representation of the previous hidden layer as input for the next hidden layer. In order to capture a higher level representation of input features, we stack another output layer on top of the DBN. This layer is used to represent the class label of an input data. Therefore, by back-propagation and having parameters initialized by the pre-training, we can optimize the weights of the deep network. This supervised optimization step is called *fine-tuning* [1].

3.2 Feature Learning

The model architecture of the DBN is shown in Fig. 1. Training of the DBN requires trial and error to determine various parameters, such as the learning rate, the momentum, the weight-cost, the initial values of the weights, the number of hidden units and the size of each mini-batch [3]. We used the MATLAB library of Tanaka *et al.* [10] to build and train the DBN. We started with the default values for all parameters as it was suggested in this library. In the first step, we heuristically searched for the number of hidden layers and the number of neurons in each layer, within the layer-wise greedy training framework [1], so that lowest errors with the default library parameters were obtained in the the training data, D_1. Next, we followed the guidelines given by Hinton [3] to adjust the learning rate, the momentum and the weight-cost to achieve lower error in the training data. The pre-trained DBN was finalized with 100 passes (epochs), a fixed learning rate of 0.001 and batch size of 5, and has a layer of real-valued visible units of dimension 50, and four hidden layers, consisting of 100, 50, 45 and 6 hidden units. The last hidden layer produces the latent features. To further reduce the error, we then proceeded to a fine-tuning step using the same training set. We ran 70 epochs with a fixed learning rate of 0.01. During both pre-training and fine-tuning, a small weight-cost of 0.0002 was used and the learning was accelerated by using a momentum of 0.9. To mitigate the risk of over-fitting the model during training, all the parameters stated above were only tuned based on data set D_1. The classification performance reported in the paper is only associated with the independent data set D_2.

3.3 Classification

We use the latent features as input for the SVM classifier[1] with the Radial Basis Function (RBF) kernel in the final stage. The parameters of SVM were determined using a grid-search algorithm which uses leave-one-core-out cross-validation on latent features obtained from D_1. In addition to the binary output of the classifier, we estimate the likelihood of the ROIs to be cancerous by using the approach presented in [12]. Finally, the classification model developed using data set D_1, was validated with the data set D_2 in the testing step (See Fig. 1).

4 Results

In order to identify the consistency of learnt features in the training set D_1, we follow a cross-validation strategy. In leave-one-core-out cross-validation experiments, an area under Receiver Operating Characteristics (ROC) curve of 0.91 is obtained for cancer prediction in the biopsy cores. The result of validation of our classification framework on data set D_2 is shown in Fig. 2. Using a threshold of 55% for the model output, 22 out of 23 benign cores can be labeled correctly. Furthermore, all cancerous cores can be identified correctly. Finally, we compare the model output against the grade assigned to each biopsy core in MRI. As shown in Fig. 2(b), the majority of MRI grades are moderate and our method can detect all of cores with moderate MRI grade correctly. Moreover, except one benign case with low MRI grade, the model correctly predicts the label of all High MRI grade biopsy cores.

We also performed an additional experiment by permuting the training and testing data. To create new training and test sets, in each permutation, we exchanged a randomly selected cancerous core larger than 1.5 mm between D1 and D2. Similarly, we exchanged a randomly selected benign core between the two data sets. This resulted in five different permutations given the distribution of large cancer cores in our data. One average, we achieved accuracy, sensitivity, and specificity of 93%, 98%, and 90%, respectively. In four out of five permutations, all cancer cores were correctly identified.

5 Discussion and Conclusion

We utilized a machine learning framework for analysing temporal ultrasound data to accurately predict prostate cancer in MRI-TRUS guided targeted prostate biopsy. A deep learning approach was used for automatic feature selection and learning of the high-level features from temporal ultrasound signals. We automatically reduced the dimensionality of the temporal ultrasound features using a DBN followed by supervised fine-tuning. Then, we used an SVM classifier along with the activation of the trained DBN to characterize the prostate tissue. The proposed approach systematically learned discriminant features from high-dimensional temporal ultrasound features to detect prostate cancer.

[1] A Library for SVMs: http://www.csie.ntu.edu.tw/∼cjlin/libsvm/

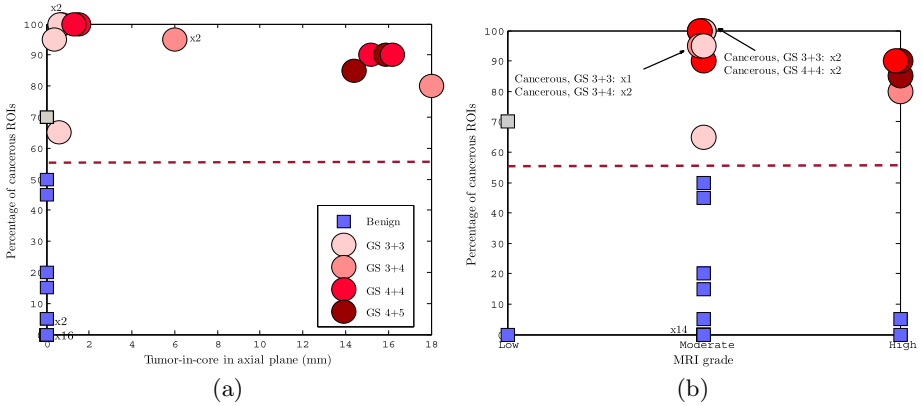

Fig. 2. Percentage of cancerous ROIs with a likelihood of more than 50% in the model output in each core: (a) The model output versus tumor-in-core-length for the validation data. (b) The model output versus MRI grade for the validation data. Cancer cores and benign cores are shown as circles and squares, respectively.

While our results with DBN are highly promising, DBN is a computationally expensive method that performs best on very large data sets. The greedy training algorithm [1] ensures that the parameters of the model can be tuned layer-by-layer, reducing the need for large sample sizes to a certain extent. We compare our approach with standard SVM and random forests implementations using the same input. In our data set, random forests outperforms SVM, and achieves sensitivity of 69% and specificity of 78%.

We built our model using a training data set consisting of temporal ultrasound data of 35 biopsy cores. The model was validated on an independent set of 36 biopsy cores. The results indicate that latent feature selection with DBN can improve the tissue typing capabilities of temporal ultrasound data analysis, which can potentially complement mp-MRI for TRUS-guided biopsy. Specifically, since all patients in our study were originally labeled as suspicious for cancer in mp-MRI, our results indicate that with a subsequent analysis of temporal ultrasound data, it might be possible to further improve the differentiation of benign and cancerous cases, hence reducing the need for unnecessary biopsy. This can be a powerful tool for surveillance of patients.

Acknowledgments. The authors would like to acknowledge the help of Dr. Jochen Kruecker and Dr. Pingkun Yan for assisting with the experiments. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR).

References

1. Bengio, Y., Lamblin, P., Popovici, D., et al.: Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems 19, 153 (2007)
2. 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 60(2), 310–320 (2013)
3. Hinton, G.: A practical guide to training RBM. Momentum 9(1), 926 (2010)
4. Imani, F., et al.: Ultrasound-based characterization of prostate cancer: An *in vivo* clinical feasibility study. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 279–286. Springer, Heidelberg (2013)
5. Khojaste, A., Imani, F., Moradi, M., Berman, D., et al.: Characterization of aggressive prostate cancer using ultrasound RF time series. SPIE Medical Imaging (2015)
6. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: A unified deep learning framework for automatic prostate MR segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 254–261. Springer, Heidelberg (2013)
7. Margel, D., Yap, S.A., Lawrentschuk, N., Klotz, L., et al.: Impact of multiparametric endorectal coil prostate magnetic resonance imaging on disease reclassification among active surveillance candidates: a prospective cohort study. The Journal of Urology 187(4), 1247–1252 (2012)
8. Moradi, M., Mousavi, P., Boag, A., Sauerbrei, E.E., Siemens, D., Abolmaesumi, P.: Augmenting detection of prostate cancer in transrectal ultrasound images using SVM and RF time series. IEEE Transactions on Biomedical Engineering 56(9), 2214–2224 (2009)
9. de Rooij, M., Hamoen, E.H., Fütterer, J.J., Barentsz, J.O., Rovers, M.M.: Accuracy of multiparametric MRI for prostate cancer detection: a meta-analysis. American Journal of Roentgenology 202(2), 343–351 (2014)
10. Tanaka, M., Okutomi, M.: A novel inference of a restricted boltzmann machine. In: International Conference on Pattern Recognition, pp. 1526–1531. IEEE (2014)
11. van Tulder, G., de Bruijne, M.: Learning features for tissue classification with the classification restricted boltzmann machine. In: Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Zhang, S., Cai, W(T.), Metaxas, D. (eds.) MCV 2014. LNCS, vol. 8848, pp. 47–58. Springer, Heidelberg (2014)
12. Uniyal, N., et al.: Ultrasound-based predication of prostate cancer in MRI-guided biopsy. In: Linguraru, M.G., Laura, C.O., Shekhar, R., Wesarg, S., Ballester, M.Á.G., Drechsler, K., Sato, Y., Erdt, M. (eds.) CLIP 2014. LNCS, vol. 8680, pp. 142–150. Springer, Heidelberg (2017)