Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H MRSI

Lukasz Matulewicz, PhD[1,2], **Jacobus F.A. Jansen, PhD**[3], **Louisa Bokacheva, PhD**[1], **Hebert Alberto Vargas, MD**[4], **Oguz Akin, MD**[4], **Samson W. Fine, MD**[5], **Amita Shukla-Dave, PhD**[1], **James A. Eastham, MD, FACS**[6], **Hedvig Hricak, MD, PhD**[4], **Jason A. Koutcher, MD, PhD**[1,4,7], and **Kristen L. Zakian, PhD**[1]

[1]Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY, United States [2]Department of Radiotherapy and Brachytherapy Planning, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice, Poland [3]Department of Radiology, Maastricht University Medical Center, Maastricht, Netherlands [4]Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY, United States [5]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY, United States [6]Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY, United States [7]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, United States

Introduction

Proton magnetic resonance spectroscopic imaging (MRSI) is a non-invasive technique that allows *in vivo* detection and quantification of metabolites within biologic tissues and has been shown to improve the specificity of conventional MRI in the detection of prostate cancer (1). Despite its potential benefits, MRSI has not gained acceptance in routine clinical practice due to a variety of factors including the length and complexity of data acquisition, processing, and analysis. The current work addresses the lack of automatic tools for analyzing the large datasets generated by prostate MRSI. Visual interpretation of the spectra by a trained spectroscopist is time consuming and requires accurate knowledge of prostate anatomy. The visual analysis relies on such features of the prostate spectra as the relative height of the choline, creatine and citrate peaks. However, it has been shown that the spectral patterns of cancer are different between the peripheral and transition zones of the prostate and thus the selection rules for cancerous voxels must take into account the location of the voxel within the prostate gland (2). Furthermore, periurethral tissue may exhibit high levels of choline-containing compounds due to the presence of glycerophosphocholine (GPC), a normal constituent of seminal fluid which is present primarily within the seminal vesicles and ejaculatory ducts but also to a lesser degree in the prostatic urethra (Fig. 1). GPC is virtually indistinguishable in vivo from phosphocholine, an important marker for prostate cancer (3). A trained spectroscopist with an expertise in zonal anatomy of the

Corresponding author: Lukasz Matulewicz, PhD, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Department of Radiotherapy and Brachytherapy Planning, Gliwice, Poland, 15 Wybrzeze AK street, 44-101 Gliwice, Poland, tel. +48 795557403, fax. +48 322313512, lukasz.matulewicz@io.gliwice.pl.

prostate is able to incorporate the information about the voxel location in the analysis of MRSI data. Therefore, a method for automated analysis of the prostate spectra that takes into account the anatomical information may be more accurate than a method relying on spectra alone.

Traditional methods of analyzing MRSI data involve calculating individual metabolite ratios obtained from areas of spectral peaks (4). In contrast, when data are analyzed with a pattern recognition technique, all information in the spectra can be used as input simultaneously, and classification of voxels can be fully automatic. Pattern recognition may be superior to partial analysis of peaks and ratios within the spectra, especially when the pattern in the data is subtle (5,6). An artificial neural network (ANN) (7,8) is a tool for pattern recognition that can rapidly process a large amount of data and has excellent generalization capability for noisy or incomplete data. ANNs have been applied to MRSI of the central nervous system and have proven useful for evaluating patients with epilepsy (9), Parkinson's disease (10) and brain cancers (11). Thus, the goal of this study is to assess the feasibility of using an ANN to automatically detect cancerous voxels from prostate MRSI datasets, and to evaluate the impact of additional information concerning the prostate's zonal anatomy on the performance of an ANN in this setting.

Patients and Methods

Patients

This retrospective study complied with the Health Insurance Portability and Accountability Act and was approved by the Institutional Review Board, with a waiver of informed consent. Eighteen patients with biopsy-proven prostate cancer, who underwent endorectal MRI/MRSI at our institution followed by radical prostatectomy with whole-mount step-section histopathology and had at least one tumor voxel detected by MRSI and confirmed by histopathology. The details of the spatial registration between MRSI and histopathology are given in section MRSI-histopathology correlation. The patient characteristics are summarized in Table 1.

MRI/MRSI data acquisition and analysis

MRI/MRSI examinations were performed on a 1.5 T whole-body unit (Signa Horizon; GE Medical Systems; Milwaukee, WI) with a body coil for excitation and a combination of a phased-array and endorectal coil (Medrad, Pittsburgh, PA) for signal reception. Anatomical images of the prostate were acquired with an axial T2-weighted fast spin-echo sequence with the following parameters: TR = 4000 ms, TE=102 ms; echo train length, 12; slice thickness, 3 mm; interslice gap, 0 mm; field of view, 14×14 cm^2; typical number of slices, 8–12; matrix, 256×192. The PROSE acquisition package (GE Medical Systems) was used to obtain MRSI data at 6.9×6.9 mm^2 in-plane resolution (Fig. 2a). The MRSI acquisition parameters were: PRESS voxel selection, TR = 1000 ms, TE = 130 ms; one average; spectral width, 1250 Hz; number of points, 512; field of view, $11.0 \times 5.5 \times 5.5$ cm^3; and $16 \times 8 \times 8$ phase encoding steps. Spectral/spatial pulses were used for water and lipid suppression within the PRESS selected volume and very selective outer voxel suppression pulses were used to reduce the contamination from surrounding tissues.

Spectral data were transferred to an Advantage workstation (GE Medical Systems), automatically processed by Functool software (GE Medical Systems) and interpreted by an experienced spectroscopist (X.X., >10 yrs of experience). In each patient, all voxels within the PRESS excitation volume were labeled as cancerous or non-cancerous according to established decision rules which incorporated both the level of polyamines relative to total choline and the ratio of choline plus creatine to citrate (12). The pertinent peaks were located at the following chemical shifts: total choline (Cho) at 3.2 ppm, creatine/ phosphocreatine (Cr) at 3.0 ppm, polyamines (PA) at 3.1 ppm and citrate (Cit) at 2.6 ppm (Fig. 3). A total of 5308 voxels were analyzed in 18 patients and 148 voxels were marked as suspicious on MRSI. The locations of the suspicious voxels were recorded on the corresponding underlying T2-weighted images.

Histopathological analysis

Prostatectomy specimens were first inked with tattoo dye (green dye on right, blue dye on left) for orientation and fixed in 10% formalin for 36 hours. The distal 5 mm portion of the apex was amputated and coned. The remainder of the gland was serially sectioned apex to base at 3-4 mm intervals and entirely submitted for paraffin embedding as whole-mounts. The seminal vesicles were amputated and submitted separately. After paraffin embedding, microsections were placed on glass slides and stained with hematoxylin and eosin. Cancer foci were outlined by an experienced pathologist (X.X., >5 yrs of experience in urologic pathology) in ink on whole-mount, apical, and seminal vesicle sections so as to be grossly visible, and photographed. These constituted the tumor maps.

MRSI-histopathology correlation

Spatial registration of the MRI/MRSI data with digitized pathologic tumor maps was performed jointly by an experienced radiologist (X.X., >5 yrs experience reading prostate MRI) and an experienced pathologist (X.X., >5 yrs of experience in urologic pathology) working in consensus. The T2-weighted images with spectral grids overlaid (see Fig. 2) were viewed side-by-side with the pathologic tumor maps. The most closely corresponding axial T2-weighted images and pathology step sections were paired based on the following anatomic landmarks: urinary bladder and seminal vesicles in superior slices, slice with largest diameter, ejaculatory duct entrance to verumontanum, thickness of the peripheral zone and position of the pseudocapsule, presence, size, and shape of the transition zone (4). Once the MRI and pathologic sections were registered, it was determined whether voxels that were classified as cancerous by MRSI corresponded to tumors in matching pathological sections. Because unavoidable gland deformation and shrinkage as well as slight differences in the angle of section limited the precision of MRI/MRSI-pathology registration, a finding of tumor in the same prostate sextant on both MRI/MRSI and pathologic tumor maps was considered a match. Voxels that were first identified by a spectroscopist as cancerous *and* confirmed by matching with histopathological maps were used as the target classification, or "gold standard", for the ANN. Of the 148 voxels that were marked suspicious by the spectroscopist, 129 voxels were labeled as true cancer voxels by comparison with histology.

Anatomical segmentation of the prostate on MRI was performed manually by a radiologist (X.X., radiology fellow trained and supervised by X.X.), who outlined the peripheral zone

(PZ), transition zone (TZ) and periurethral region on the T2-weighted images and the superimposed spectroscopy matrix (Fig. 2b). The percentage of the area within each voxel located in the PZ, TZ, periurethral region (U) and/or outside (O) the prostate was visually estimated (X.X.) and recorded.

Among the prostate voxels, 1139/5308 (21.5% of all voxels) contained more than 50% PZ, 1457/5308 (27.4%) contained more than 50% TZ and 7.3% contained more than 10% U.

The real parts of all spectra were exported using a custom Matlab script (Mathworks; Natick, MA), and the spectral range of 4.3–0.4 ppm (256 points) was analyzed.

Artificial Neural Network Analysis

The purpose of the ANN was to identify voxels which had been designated by the spectroscopist as cancer and confirmed by histopathology. ANN analysis was performed on a Pentium 4, 3.2 GHz computer using Statistica Neural Networks software (StatSoft Inc., Tulsa, OK). The input data of the ANN models consisted either of MRSI spectra (intensities at ppm positions) or spectra and additional anatomical information. The output of the ANN was a continuous confidence level value ranging between 0 and 1, which indicates the probability of a given voxel belonging to one of two possible classes, cancer or non-cancer, where confidence of 0 indicates that the voxel is certainly cancerous and 1 indicates a certainly non-cancerous voxel. The pattern recognition ability of the ANN models was characterized using the receiver operator characteristic (ROC) curve, which was created by calculating the sensitivity and specificity of the model as the confidence level separating cancer and non-cancer assumes a series of values between 0 and 1. For each trial, the full dataset (5308 spectra) was randomly split into training set (70% of all spectra), test set (15%) and validation dataset (15%). The training set was used for adjusting the network connection weights; the test set was used to monitor the progress of ANN training and prevent overtraining; finally, the model was applied to a validation data set that had not been shown to the model during the training process. The ANN model that provided the highest area under the ROC curve (AUC) was selected as the optimally performing model.

The ANN was implemented as a feed-forward multilayer perceptron (MLP) with three layers: input, hidden and output layers. The input and output layers both contained a single neuron. We decided to use only one neuron in the hidden layer because the performance of such model for the test set reached an AUC of 0.99. Increasing the number of neurons in the hidden layer may lead to overtraining and poor generalization, which occur when excessive flexibility of the model improves its performance on the test set, but compromises the model's ability to correctly evaluate future observations (13). The back propagation Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative algorithm was used to train the model (14). With this algorithm, a classification was first made and compared with the target classification and the error was back-propagated through the network to adjust the connection weights between neurons to achieve a closer match between the output and the observed data. The error was then recalculated and the process repeated on the next iteration. The error function was calculated as the cross entropy. The logistic sigmoid activation function was used between the input and the hidden layer and the softmax (normalized exponential) activation function for the output layer.

Two models were created with the same configurations of hidden and output layers, but with different input variables. Model 1 used as input only MRSI spectra (256 variables as resonance intensities at given ppm positions) and model 2 used the spectra and the anatomical segmentation information (260 variables total: the same 256 variables as in model 1 plus four additional variables as percentage of PZ, TZ, U, and O). The final models 1 and 2 were selected as the models that provided the highest AUC values for the test set among one hundred models for each model type constructed automatically by the software with the same data randomly assigned to the training, validation and test sets each time a new model was created. The pattern recognition abilities of the two models were compared by assessing their ROCs on the validation sets using the method of DeLong *et al.* (15) implemented in Analyse-it for Microsoft Excel (Analyse-it Software, Ltd., United Kingdom) and the sensitivity and specificity at the confidence level of 0.5.

To determine whether the variations of the signal-to-noise ratio (SNR) influenced the pattern recognition capabilities of the ANN models, the SNR values were defined for each voxel as the amplitude of the choline signal divided by the standard deviation of the signal in the range 0.5 – 0 ppm, which contains only noise (16). The mean SNR values for the correctly and incorrectly classified cancerous spectra were compared using the Mann-Whitey U-test.

To establish the relative importance of the input variables for the accuracy of voxel classification, the ANN model's sensitivity to each input variable (not to be confused with the sensitivity of pattern recognition) was assessed by calculating the error in response of the ANN to the change in each variable. The error was computed after the data set was repeatedly submitted to the network, with each variable replaced with its mean value determined from the training sample. The sensitivity to the input variable was expressed as the ratio of the error calculated with the variable missing to the error obtained from the complete set of variables. A variable was considered important if the model's sensitivity to this variable was greater than 1.

Results

Fig. 4 presents ROC curves from training, test and validation sets for the optimal versions of model 1 (spectra only, Fig. 4a) and model 2 (spectra plus segmentation, Fig. 4b). The AUC values for all sets classified by model 1 (training, 0.955; test, 0.976; validation, 0.949) were lower than the corresponding AUCs for model 2 (training, 0.988; test, 0.993; validation, 0.968). The difference between the AUCs of the two models was the highest for the training set.

The performance of the ANN models for the validation data set is summarized in Table 2. The model 2 yielded a significantly higher AUC than model 1 (p = 0.03) and had a narrower confidence interval. At the confidence level of 0.5, model 1 classified correctly only 50% of cancer voxels, but model 2 was able to correctly classify two more voxels and had the sensitivity of 62.5%. The difference between the specificity values of the two models was smaller, because of the small proportion of cancer voxels relative to the number of benign voxels. The disproportionately large percentage of non-cancerous voxels in the dataset (5179 out of 5308 voxels, or 97.6%) presented a challenge in developing the ANN. The

model correctly identifies the vast majority of the healthy voxels while failing to detect a large percentage of cancer voxels and yields a deceptively high the AUC, which does not correctly reflect the classification rate in the cancer class.

The capabilities of either model to correctly recognize the cancerous voxels did not appear to depend on the variations of SNR (Table 3). The mean SNR values among the correctly and incorrectly classified voxels in all three datasets combined (training, test and validation sets) were not significantly different for either model.

Fig. 5 shows the ten most important input variables identified by the ANN model's sensitivity analysis. For model 1, the most important variables were the ppm positions corresponding to the choline peak in the 3.15–3.26 ppm range with the sensitivities of about 1.2 – 1.3 (Fig. 6a). For model 2, the most important variable was the percentage of PZ, %PZ, for which the model sensitivity, equal to 2.8, was considerably higher than the sensitivity to the variables corresponding to the ppm positions corresponding to choline, which was in the range of 1.3–1.5 (Fig. 6b). The next most important segmentation variable was the percentage of region U, %U, which was in the eighteenth position with the model sensitivity of 1.19.

Discussion

We implemented and compared two models for automatic classification of MRSI voxels in the prostate gland: model 1, which used only spectra as input, and model 2, which used the spectra plus information from anatomical segmentation. The models were trained, tested and validated using spectra from voxels that the spectroscopist had designated as cancer and that were verified on histopathological maps. At ROC analysis, model 2 provided significantly better classification of voxels than did model 1. At the confidence level of 0.5, model 2 was able to correctly detect more cancerous voxels than model 1. Variations in the SNR did not influence the pattern recognition capabilities of either model; hence, the errors in classification were caused by other sources, which are discussed below.

For a predictive model, the quality of the training dataset is of paramount importance for accurate outcome. In our analysis, the quality of the training set was based on two assumptions— namely, that cancerous voxels showed a unique spectral pattern and that the spectroscopist was able to recognize *all* voxels with this spectral pattern (12). Thus, if a cancerous voxel was marked by the spectroscopist as non-cancerous either because the spectrum did not satisfy the selection criteria or because the spectroscopist missed it, that voxel was not compared with histopathology. If the first assumption is invalid, that is, cancer does not in fact have a unique spectral pattern, it is a limitation of MRSI as a method and is impossible to correct. However, in the second scenario, if some of the cancer voxels are missed by the spectroscopist, the ANN may be able to correct the spectroscopist's error and correctly classify the voxel, using knowledge gained from the voxels in the training set.

In our study, sensitivity and specificity were defined with respect to a potentially imperfect reference classification. The target classification was the result of a one-way matching of voxels detected on MRSI by the spectroscopist to the histopathological maps; however,

cancers identified only on histopathologic maps were not "back"-correlated to voxels in the MRSI.

False positives, i.e., voxels incorrectly classified as cancerous based on their spectral patterns, originate from sources such as prostatitis, benign prostatic hyperplasia, or the elevated choline peak originating from the urethra (17,18). The limited number of k-space samples in each dimension gives rise to a point-spread function (19) that can "smear" the choline signal over voxels adjacent to the urethra. Based on previous experience, the spectroscopist can often recognize this type of spectral "bleed". The ANN model 2 was sensitive to the designation of a voxel as being in the U region, with this input variable ranking eighteenth out of 260 variables. However, our model does not take into account the potential for "bleed" into voxels adjacent to the urethra. In theory, one could teach the neural network to recognize a multivoxel pattern of Cho emanating from a maximum in the urethra. However, the feasibility of this idea needs to be tested.

Sensitivity to input variables was higher for model 2 than for model 1. Apart from segmentation, for both models, the most important variables were the ppm positions around 3.21 ppm corresponding to choline. This finding is in accord with the previously determined decision criteria, which identify choline as a marker of malignancy (20). For model 2 (with segmentation), the percentage of PZ within a given voxel was the most important input variable.

In this study, we attempted to address some of the well-described theoretical disadvantages of ANNs. The first is the risk of "overtraining". As a network is being trained, initially the error of recognition drops rapidly with every iteration, but then the rate of improvement diminishes. Further training improves fitting of this particular training set, but makes the model less capable of correctly predicting any other dataset, a problem that is referred to as poor generalization of the ANN. We used two common approaches to prevent this phenomenon (13). Firstly, we used the smallest possible number of neurons in the hidden layer. In general, increasing the number of hidden units increases the modeling power of the neural network but also makes it larger, more difficult to train, slower to operate, and more prone to overfitting. Secondly, to avoid overtraining we also applied the so called "early-stopping" method, in which the training process was stopped when the error of classification of the test set ceased to decrease.

Many studies using ANN provide the results only for training and test sets (21), which may present an overly optimistic estimation of the model's performance. In our study the final models were tested with the independent validation subset (i.e., data never shown to the model during the training process), in order to ensure that the results were not artifacts of the training process. However, a weak point in this preliminary analysis is the consideration of every spectrum independently without providing information from adjacent spectra in a particular patient. The only information about anatomical localization is the PZ, TZ, U or O labeling.

Although ANNs can "learn" from specific patterns, it is very difficult to understand how exactly the model learns and how different weights are attached to variables during this

process. Thus it is difficult to ensure that clinical and methodological sense always prevails in the interpretation of results. We addressed this by performing the ANN model's sensitivity analysis, which identified specific points of the MRSI spectra that were especially important for the model in distinguishing cancerous from non-cancerous voxels. Our sensitivity analysis confirmed that choline, as determined previously by the visual method of MRSI spectral analysis (12), is an important marker for making this distinction.

In our study, spatial matching of MRSI and histopathological maps was limited to the sextant level due to the difficulties arising from variability in the size, shape and contour of the resected gland as well as angle of tilt in sectioning. In the future, development of a mechanism for fusing MRSI data with the whole-mount step-section pathology maps would be helpful for informing the ANN about the locations of cancerous voxels.

Many studies have looked at the usefulness of incorporating clinical data such as age, prostate volume, PSA level and digital rectal examination and transrectal ultrasound results into ANNs for predicting prostate cancer (22-33). However, these studies have attempted to recognize only the presence or absence of the cancer and the clinical outcome. In our study, ANNs were used for detection *and* localization of cancer based on MRSI data, and therefore our ANN model used different input variables.

A major limitation of the current study was the low number of patients included. However, our goal was to demonstrate the feasibility of including anatomical data in an ANN model as well as the potential advantage for correctly identifying cancer-containing voxels. A larger population to validate and refine the model is necessary in order to demonstrate clinical utility.

Another limitation of the current study was the manual segmentation of the prostatic zones, which is time consuming. The automatic prostate segmentation is currently an area of active study (34-36).

In conclusion, though our preliminary findings require validation in a larger prospective clinical trial, we have shown the feasibility of using an ANN to automatically analyze in-vivo MRSI and thus detect and localize prostate cancer non-invasively. The incorporation of novel biomarkers and non-invasive imaging data into predictive tools has the potential to improve clinical decision-making in patients with prostate cancer (37). The ANN model presented herein is a significant step toward an automatic, user-independent and fast method of analyzing prostate MRSI data, and could facilitate the clinical use of MRSI to boost predictive accuracy in prostate cancer assessment.

Acknowledgments

References

1. Hricak H, Choyke PL, Eberhardt SC, Leibel SA, Scardino PT. Imaging Prostate Cancer: A Multidisciplinary Perspective. Radiology. 2007; 243(1):28–53. [PubMed: 17392247]

2. Zakian KL, Eberhardt S, Hricak H, et al. Transition zone prostate cancer: metabolic characteristics at 1H MR spectroscopic imaging--initial results. Radiology. 2003; 229(1):241–247. [PubMed: 12920178]

3. Villeirs G, Oosterlinck W, Vanherreweghe E, De Meerleer G. A qualitative approach to combined magnetic resonance imaging and spectroscopy in the diagnosis of prostate cancer. European Journal of Radiology. 2010; 73(2):352–356. [PubMed: 19084363]

4. Zakian KL, Sircar K, Hricak H, et al. Correlation of proton MR spectroscopic imaging with gleason score based on step-section pathologic analysis after radical prostatectomy. Radiology. 2005; 234(3):804–814. [PubMed: 15734935]

5. El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR in Biomedicine. 1997; 10(3):99–124. [PubMed: 9408920]

6. Hagberg G. From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. NMR Biomed. 1998; 11(4-5):148–156. [PubMed: 9719569]

7. Rosenblatt, F. Principles in Neurodynamics. Washington DC: Spartan Books; 1962.

8. Widrow, B.; Hoff, M. IRE WESCON Convention Record. New York: 1960. Adaptive Switching Circuits; p. 96-104.

9. Bakken IJ, Axelson D, Kvistad KA, et al. Applications of neural network analyses to in vivo 1H magnetic resonance spectroscopy of epilepsy patients. Epilepsy Res. 1999; 35(3):245–252. [PubMed: 10413319]

10. Axelson D, Bakken IJ, Susann Gribbestad I, Ehrnholm B, Nilsen G, Aasly J. Applications of neural network analyses to in vivo 1H magnetic resonance spectroscopy of Parkinson disease patients. J Magn Reson Imaging. 2002; 16(1):13–20. [PubMed: 12112498]

11. Poptani H, Kaartinen J, Gupta RK, Niemitz M, Hiltunen Y, Kauppinen RA. Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. J Cancer Res Clin Oncol. 1999; 125(6): 343–349. [PubMed: 10363566]

12. Shukla-Dave A, Hricak H, Moskowitz C, et al. Detection of prostate cancer with MR spectroscopic imaging: an expanded paradigm incorporating polyamines. Radiology. 2007; 245(2):499–506. [PubMed: 17890357]

13. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. Semin Urol Oncol. 2002; 20(2):89–95. [PubMed: 12012294]

14. Bishop, CM. Neural networks for pattern recognition. Vol. xvii. Oxford New York: Clarendon Press; Oxford University Press; 1995. p. 482

15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44(3):837–845. [PubMed: 3203132]

16. Jiru F. Introduction to post-processing techniques. Eur J Radiol. 2008; 67(2):202–217. [PubMed: 18436403]

17. Shukla-Dave A, Hricak H, Eberhardt SC, et al. Chronic prostatitis: MR imaging and 1H MR spectroscopic imaging findings--initial observations. Radiology. 2004; 231(3):717–724. [PubMed: 15163811]

18. Verma S, Rajesh A, Ftterer J, et al. Prostate MRI and 3D MR spectroscopy: how we do it. AJR, American journal of roentgenology. 2010; 194(6):1414–1426. [PubMed: 20489079]

19. Wu CL, Jordan K, Ratai E, et al. Metabolomic imaging for human prostate cancer detection. Science translational medicine. 2010; 2(16):16ra18–16ra18.

20. Zakian K, Shukla-Dave A, Ackerstaff E, Hricak H, Koutcher J. 1H magnetic resonance spectroscopy of prostate cancer: biomarkers for tumor characterization. Cancer biomarkers. 2008; 4(4-5):263–276. [PubMed: 18957715]

21. Lisboa P, Taktak AFG. The use of artificial neural networks in decision support in cancer: a systematic review. Neural Networks. 2006; 19(4):408–415. [PubMed: 16483741]

22. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. J Clin Oncol. 2002; 20(4):921–929. [PubMed: 11844812]

23. Chun FK, Haese A, Ahyai SA, et al. Critical assessment of tools to predict clinically insignificant prostate cancer at radical prostatectomy in contemporary men. Cancer. 2008; 113(4):701–709. [PubMed: 18553365]

24. Stephan C, Cammann H, Bender M, et al. Internal validation of an artificial neural network for prostate biopsy outcome. Int J Urol. 2010; 17(1):62–68. [PubMed: 19925616]

25. Kawakami S, Numao N, Okubo Y, et al. Development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. Eur Urol. 2008; 54(3):601–611. [PubMed: 18207312]

26. Lee HJ, Hwang SI, Han SM, et al. Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: comparison of multiple logistic regression, artificial neural network, and support vector machine. Eur Radiol. 2010; 20(6):1476–1484. [PubMed: 20016902]

27. Meijer RP, Gemen EF, van Onna IE, van der Linden JC, Beerlage HP, Kusters GC. The value of an artificial neural network in the decision-making for prostate biopsies. World J Urol. 2009; 27(5):593–598. [PubMed: 19562346]

28. Ecke TH, Bartel P, Hallmann S, et al. Outcome prediction for prostate cancer detection rate with artificial neural network (ANN) in daily routine. Urol Oncol. 2010

29. Lawrentschuk N, Lockwood G, Davies S, et al. Predicting prostate biopsy outcome: artificial neural networks and polychotomous regression are equivalent models. Int Urol Nephrol. 2010

30. Finne P, Finne R, Auvinen A, et al. Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology. 2000; 56(3):418–422. [PubMed: 10962306]

31. Porter CR, O'Donnell C, Crawford ED, et al. Predicting the outcome of prostate biopsy in a racially diverse population: a prospective study. Urology. 2002; 60(5):831–835. [PubMed: 12429310]

32. Remzi M, Anagnostou T, Ravery V, et al. An artificial neural network to predict the outcome of repeat prostate biopsies. Urology. 2003; 62(3):456–460. [PubMed: 12946746]

33. Porter CR, Crawford ED. Combining artificial neural networks and transrectal ultrasound in the diagnosis of prostate cancer. Oncology (Williston Park). 2003; 17(10):1395–1399. discussion 1399, 1403-1396. [PubMed: 14606364]

34. Toth R, Ribault J, Gentile J, Sperling D, Madabhushi A. Simultaneous Segmentation of Prostatic Zones Using Active Appearance Models With Multiple Coupled Levelsets. Computer vision and image understanding: CVIU. 2013; 117(9):1051–1060. [PubMed: 23997571]

35. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. Medical image analysis. 2013; 17(2):219–235. [PubMed: 23294985]

36. Viswanath SE, Bloch NB, Chappelow JC, et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. J Magn Reson Imaging. 2012; 36(1):213–224. [PubMed: 22337003]

37. Shariat S, Kattan M, Vickers A, Karakiewicz P, Scardino P. Critical review of prostate cancer predictive tools. Future Oncology. 2009; 5(10):1555–1584. [PubMed: 20001796]

38. Spencer NG, Eykyn TR, deSouza NM, Payne GS. The effect of experimental conditions on the detection of spermine in cell extracts and tissues. NMR Biomed. 2010; 23(2):163–169. [PubMed: 19757478]
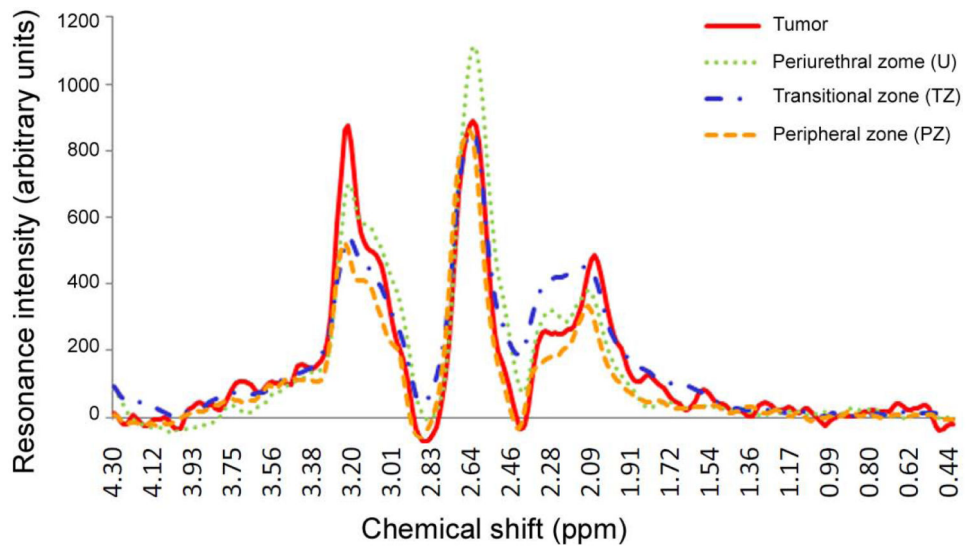
Fig. 1.
Averaged spectra from different prostate zones. The curves were derived from 129 cancerous spectra; 1139 peripheral zone spectra (> 50% peripheral zone within the voxel); 1457 transition zone spectra (> 50% transition zone within the voxel); and 389 periurethral zone spectra (>= 10% periurethral tissue within the voxel). While the most characteristic marker of prostate cancer on MRSI is an elevated choline signal at 3.21 ppm, elevated choline is also observed in the normal periurethral region. The averaged spectra illustrate the difficulty of correctly classifying the voxels close to the urethra and suggest that anatomical segmentation may be helpful for distinguishing the periurethral spectra from cancerous spectra. Cancer tissue spectra also reveal a relatively elevated peak at 2.06 ppm, which is likely to be spermine (38); however, this region is in the transition band of the spectral-spatial refocusing pulses (4).
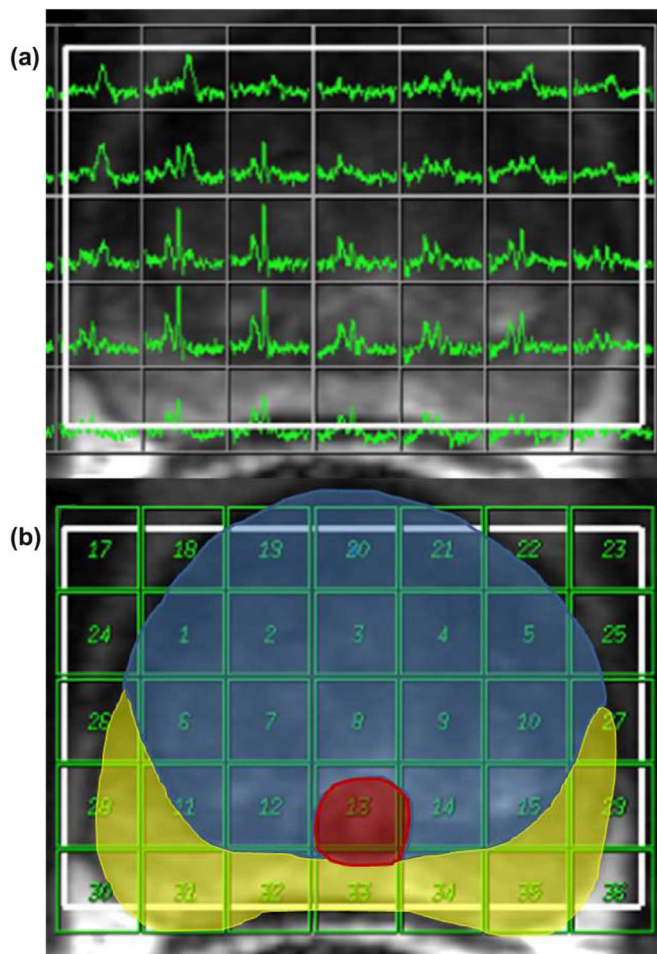
Fig. 2.
(a) A grid of proton MR spectra overlaid on a T2-weighted image of the prostate. The spectroscopic acquisition parameters were: PRESS voxel selection, TR = 1000 ms, TE = 130 ms; one average; spectral width, 1250 Hz; number of points, 512; field of view, $11.0 \times 5.5 \times 5.5$ cm^3; and $16 \times 8 \times 8$ phase encoding steps. (b) Example of zonal segmentation of the prostate on T2-weighted image with overlaid MRSI grid. Prostatic zones are indicated by color overlays: peripheral zone (yellow), transition zone (blue), periurethral tissue (red). The numeral in each voxel in (b) is assigned by the manufacturer's software and has no meaning for the analysis.
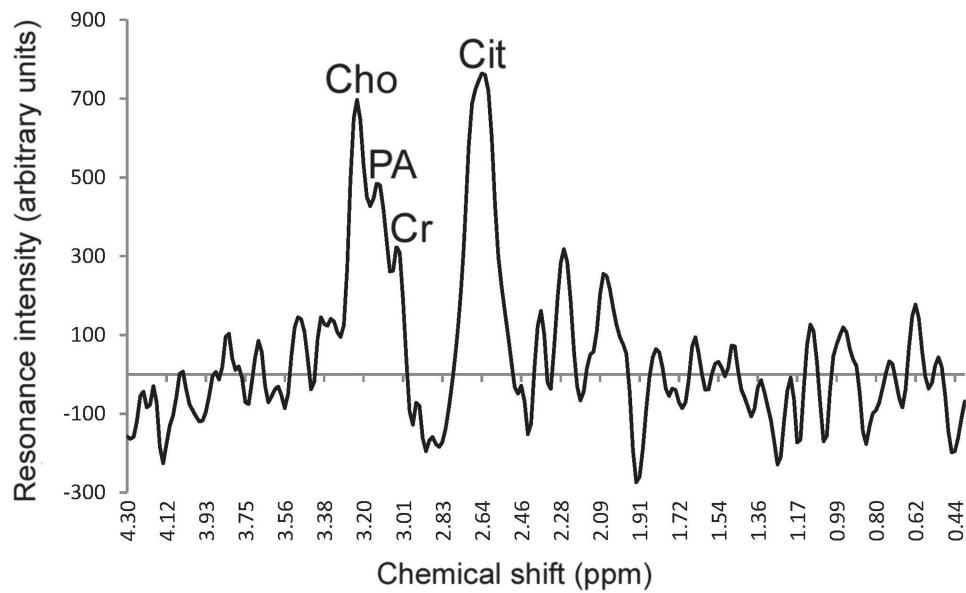
Fig. 3.
A representative 1.5 T prostate cancer voxel spectrum from peripheral zone (256 spectral points) shows typically elevated Cho relative to Cr and Cit peaks. Abbreviations: choline, Cho; polyamines, PA; creatine, Cr; and citrate, Cit. The voxel is suspicious for cancer based on the criteria published in reference 12.
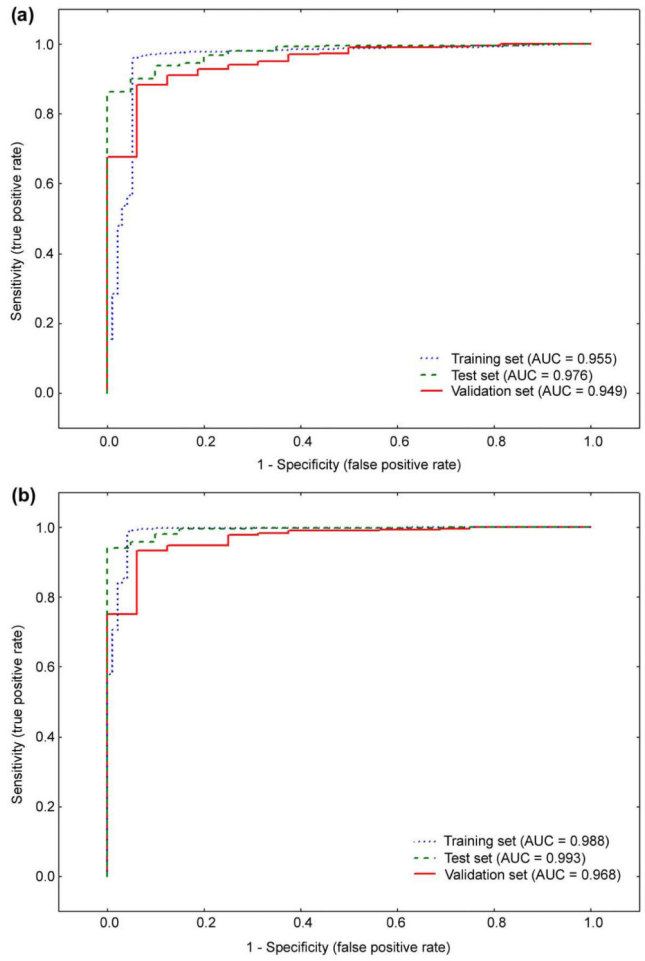
Fig. 4.

Receiver operating characteristic curves of the pattern recognition models for training, test and validation sets: (a) model 1, with only spectra as input; (b) model 2, using spectra and anatomical segmentation. The optimally performing models 1 and 2 were chosen out of one hundred automatically generated models of each kind, with the full dataset split into training, test and validation sets every time a new model is created. The legends indicate AUC for each dataset. Model 2 (b) has higher AUCs than model 1 (a) for each of the three datasets.
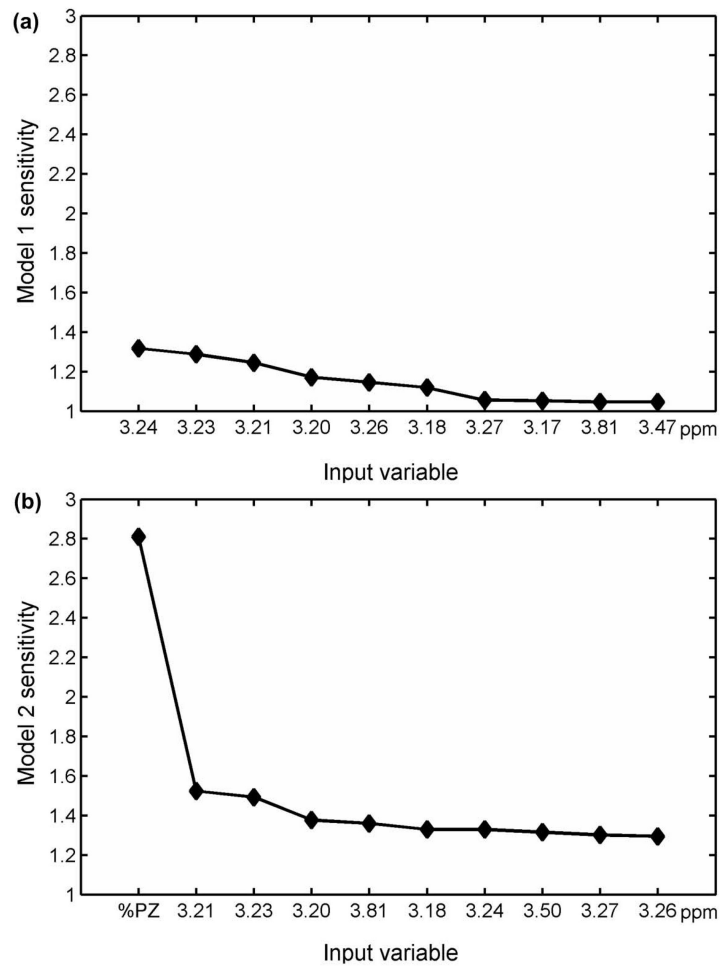
Fig. 5.
ANN model's sensitivity for the top ten most important variables, in the order of descending relative importance from left to right: (a) model 1, using spectra only; (b) model 2, using spectra and anatomical segmentation. For model 1 (a), all top ten variables originate from the Cho ppm range. For model 2 (b), the most important variable is %PZ followed by the Cho ppm intensities, which are similar to the Cho ppm positions for model 1. The overall sensitivity of model 2 is higher than the sensitivity of model 1.

Table 1

Patient characteristics.

Parameter		Value
Number of patients		18
Age at MRI [years], median (range)		55 (36–71)
Serum prostate-specific antigen [ng/mL], median (range)		4.63 (0.58– 11.6)
Clinical stage, number (%)		
	1c	14 (78)
	2a	4 (22)
Time between MRI/MRSI and prostatectomy [days], median (range)		63 (3–148)
Biopsy Gleason score, number (%)		
	3+3	12 (67)
	3+4	4 (22)
	4+3	2 (11)
	4+4	0
Prostatectomy Gleason score, number (%)		
	3+3	6 (33)
	3+4	9 (50)
	4+3	2 (11)
	4+4	1 (6)

Table 2

Performance of the ANN model 1 (spectra only) and model 2 (spectra plus anatomical segmentation) in validation data set (796 voxels): area under ROC curve (AUC) and sensitivity and specificity at confidence level of 0.5.

	AUC		Sensitivity (%)	Specificity (%)
	Estimate	95% CI		
Model 1	0.949	0.909 – 0.990	50.0 (8/16)	98.7 (770/780)
Model 2	0.968	0.937 – 0.999	62.5 (10/16)	99.0 (772/780)

Note: AUC for model 2 is significantly higher than AUC of model 1 (p = 0.03) (15).

Table 3

Signal-to-noise ratios (SNRs) of spectra correctly and incorrectly classified as cancer by model 1 and model 2 at confidence threshold 0.5 for training, test and validation data sets combined.

Model	Classification correctness	Number of voxels	SNR [mean±SD]	p-value[*]
Model 1	correct	48	5.5 ± 2.7	0.813
	incorrect	81	5.6 ± 2.3	
Model 2	correct	107	5.5 ± 2.5	0.819
	incorrect	22	5.8 ± 2.3	

[*] Mann-Whitney U test for SNR of correctly classified voxels versus incorrectly classified voxels for a given model.