Technology in Cancer Research & Treatment
Volume 17: 1-11

DOI: 10.1177/1533034618775530
journals.sagepub.com/home/tct
SAGE

Deep Learning Role in Early Diagnosis of Prostate Cancer

Islam Reda, MSc[1,2], Ashraf Khalil, PhD[3], Mohammed Elmogy, PhD[1,2], Ahmed Abou El-Fetouh, PhD[1], Ahmed Shalaby, PhD[2], Mohamed Abou El-Ghar, PhD[4], Adel Elmaghraby, PhD[5], Mohammed Ghazal, PhD[3], and Ayman El-Baz, PhD[2]

Abstract

The objective of this work is to develop a computer-aided diagnostic system for early diagnosis of prostate cancer. The presented system integrates both clinical biomarkers (prostate-specific antigen) and extracted features from diffusion-weighted magnetic resonance imaging collected at multiple *b* values. The presented system performs 3 major processing steps. First, prostate delineation using a hybrid approach that combines a level-set model with nonnegative matrix factorization. Second, estimation and normalization of diffusion parameters, which are the apparent diffusion coefficients of the delineated prostate volumes at different *b* values followed by refinement of those apparent diffusion coefficients using a generalized Gaussian Markov random field model. Then, construction of the cumulative distribution functions of the processed apparent diffusion coefficients at multiple *b* values. In parallel, a K-nearest neighbor classifier is employed to transform the prostate-specific antigen results into diagnostic probabilities. Finally, those prostate-specific antigen–based probabilities are integrated with the initial diagnostic probabilities obtained using stacked nonnegativity constraint sparse autoencoders that employ apparent diffusion coefficient–cumulative distribution functions for better diagnostic accuracy. Experiments conducted on 18 diffusion-weighted magnetic resonance imaging data sets achieved 94.4% diagnosis accuracy (sensitivity = 88.9% and specificity = 100%), which indicate the promising results of the presented computer-aided diagnostic system.

Keywords

prostate cancer, CAD, PSA, ADC, SNCSAE

Abbreviations

ADC, apparent diffusion coefficient; AUC, area under the curve; CAD, computer-aided diagnostic; CDF, cumulative distribution function; CG, central gland; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; DRE, digital rectal examination; DW-MRI, diffusion-weighted magnetic resonance imaging; GGMRF, generalized Gaussian Markov random field; KNN, K-nearest neighbor; LOSO, leave-one-subject-out; NMF, nonnegative matrix factorization; PSA, prostate-specific antigen; PZ, peripheral zone; ROC, receiver operating characteristic; SNCSAE, stacked nonnegativity constraint sparse autoencoders; TRUS, transrectal ultrasound.

Received: October 09, 2017; Revised: February 16, 2018; Accepted: February 26, 2018.

Introduction

Prostate cancer is one of the most frequent cancers identified among the male population in the United States and accounts for the highest mortality rate due to cancer second only to lung cancer. According to statistics published by the American Cancer Society in 2016, around 180 890 new patients were diagnosed and around 26 120 patients passed away due to prostate cancer.[1] The incidence of prostate cancer increases as age progresses. By 2030, it is estimated there will be up to

[1] Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
[2] Department of Bioengineering, University of Louisville, Louisville, KY, USA
[3] Electrical and Computer Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates
[4] Radiology Department, Mansoura University, Mansoura, Egypt
[5] Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY, USA

Corresponding Author:
Islam Reda, MSc, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt.
Email: islamcis@yahoo.com

1 700 000 prostate cancer incidences worldwide, and the related number of annual deaths will be around 500 000.[2] Fortunately, the sooner prostate cancer is detected, the more likely it is treated, and the mortality rate is minimized.

Current screening techniques of prostate cancer include digital rectal examination (DRE),[3] prostate-specific antigen (PSA) blood test,[4] and needle biopsy.[5] Each of these techniques have their own shortcomings. In the DRE test, a doctor examines the prostate manually to identify anomalies in volume or hardness. The DRE cost is relatively low, but it is highly invasive. Some peripheral zone tumors can be identified using the DRE. However, most of the central zone and transitional zone tumors, as well as tumors too small to be palpated, cannot be detected through the DRE. As a result, the positive predictive value, sensitivity, and specificity of the DRE are low.[3] The most prevailing prostate examination measures the PSA enzyme concentration in the blood. An increased PSA level higher than 4 ng/mL (nanograms per milliliter) might likely indicate prostate cancer. However, the elevated levels may also be due to other reasons, such as prostatitis or hyperplasia. In general, the positive predictive value, sensitivity, and specificity of the PSA screening are better than the DRE test.[3] However, both DRE and PSA are indicators of prostate cancer and are not conclusive diagnosis techniques.

If either the DRE or PSA test raises suspicion, patients undergo further testing, such as needle biopsy, to confirm the existence or nonexistence of the cancer. Transrectal ultrasound (TRUS)-guided biopsy acquires small tissue specimens from the prostate gland for evaluation by a pathologist. The Gleason grading system is the standard method used by pathologists for visual assessment of acquired specimens. The Gleason grading system is based on evaluating the 2 most predominant tumor patterns in the acquired specimen. A pathologist evaluates each pattern on a scale from 1 to 5, where 5 represents the most aggressive tumor. The Gleason score is the result of summing the scores of these 2 patterns. A score of 6 or more indicates the presence of prostate cancer. However, there is a possibility of missing a cancer tumor due to the small number of biopsy specimens, random nature of sampling, and poor resolution of TRUS. Magnetic resonance imaging (MRI)/US-guided biopsy has been demonstrated to perform better than TRUS-guided biopsy. Even though biopsy is the most precise technique for detecting cancer, it is highly invasive, expensive, and a painful tool for detecting prostate cancer and determining its aggressiveness. Because of these shortcomings, accurate, sensitive, specific, and noninvasive diagnostic techniques are in a high demand.

Today's computer-aided diagnostic (CAD) systems analyze images from various modalities, such as, ultrasound and MRI, to detect and localize prostate cancer, as well as evaluate its size and extent. In clinical applications, each of these modalities have pros and cons. The most prevalent prostate imaging modality is TRUS as it is used to guide the needle biopsy and estimate the prostate volume.[6] In comparison to other imaging modalities, the TRUS is portable, inexpensive, and generates real-time images. The negative aspects of TRUS imagery are low contrast, noisy due to small signal-to-noise ratio, the existence of speckles, and shadow artifacts.[7] The interpretation of the TRUS images is highly affected by the subjective nature of the examination and the experience of the clinician. Therefore, it is hard to precisely detect tumors and/or identify the cancer stage with the TRUS images.

Various MRI modalities have been used in CAD systems for prostate cancer diagnosis. Diffusion-weighted magnetic resonance imaging (DW-MRI) is the most recent MRI modality for diagnosing prostate cancer. Diffusion-weighted magnetic resonance imaging employs the diffusion of water molecules to indirectly identify cellularity of tissues. Cancerous prostate regions are characterized by increased cell densities, which result in more constrained diffusion compared to healthy tissues. Even though the contrast of DW-MR images is not as good as the contrast of dynamic contrast-enhanced MRI (DCE-MRI), the acquisition time of DW-MRI data is much shorter, and it does not involve the use of any contrast agents.[8,9] In general, using DW-MRI for diagnosing prostate cancer results in higher accuracy than using DCE-MRI or T2-weighted MRI.[10]

Viswanath *et al*[11] introduced a CAD system for detecting prostate cancer in both the central gland (CG) and the peripheral zone (PZ) from T2-weighted MRI. In their system, 110 textural features were extracted. Then, a feature selection approach was performed to choose the minimum number of features with the best accuracy, in terms of the area under the curve (AUC), using a quadratic discriminant analysis classifier for both the CG and the PZ. Their experiments on a data set of 22 participants showed that applying feature selection resulted in a better accuracy than using the whole set of textural feature. The resulting AUC under the receiver operating characteristic (ROC) curve were 0.86 and 0.73 for CG cancer and PZ cancer, respectively. Hambrock *et al*[12] presented a CAD system to help radiologists in differentiating malignant lesions from benign ones in both the TZ and the PZ. Two linear discriminant analysis classifiers, one for the PZ and the other for the TZ, were used to estimate malignancy likelihood using apparent diffusion coefficient (ADC) maps and DCE-based features extracted from regions of interest. The experimental results on a data set of 34 patients showed that the use of their CAD system raises the accuracy of the diagnosis for less-experienced radiologists in terms of AUC from 0.81 to 0.91, which is equal to the accuracy for experienced radiologists. Litjens *et al*[13] proposed a multiparametric CAD system for prostate cancer diagnosis from DW-MRI, T2-weighted MRI, DCE-MRI, and proton density-weighted MRI. The diagnosis is performed in 2 steps: first, initial candidate regions are detected. Then, those regions are classified to generate cancer likelihood maps. For classification, the random forest (RF) classifier was used and an AUC of 0.91 was achieved. Kwak *et al*[14] used DW-MRI at a high *b* value with T2-weighted MRI in their proposed CAD system. They utilized a textural feature selection approach of 3 stages to obtain the features with the best discriminative capabilities. The resulting AUC was 0.89. Peng *et al*[15] evaluated the correlation between a number of features extracted from

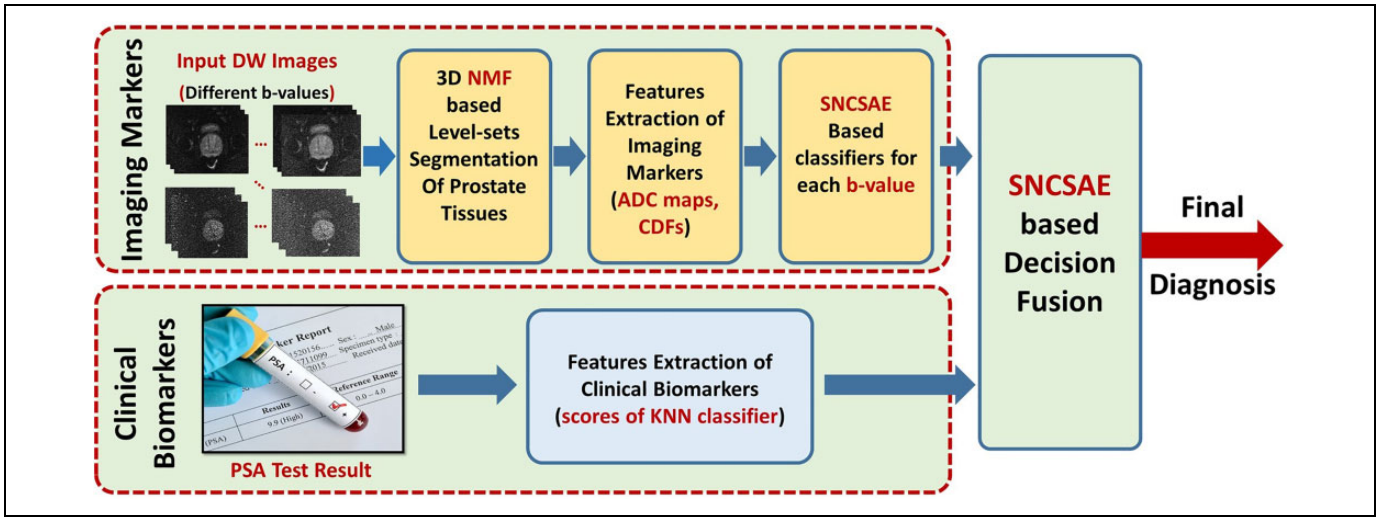

Figure 1. Framework of the presented computer-aided diagnostic (CAD) system for prostate cancer diagnosis.

multiparametric MRI and the aggressiveness of the cancer. Their experiments demonstrated a correlation between ADC values and Gleason score. This correlation was confirmed in a recent study.[16]

There are 2 main limitations of the developed CAD systems for prostate cancer diagnosis from DW-MRI. First, according to the literature, most of these CAD systems use only images acquired at a certain b value. The used b values are not consistent and differ from a system to another. Therefore, there is no agreement on which b value will provide the best diagnosis. Second, most investigators have just used imaging markers and they have not integrated them with the current clinical biomarkers which affect the final accuracy. To overcome these limitations, we are the first group who investigate the integration of imaging markers with clinical biomarkers to develop an accurate and robust system for early diagnosis of prostate cancer. Moreover, our presented CAD system uses DW-MRI data collected at multiple b values. Therefore, it is not sensitive to the selection of a b value. One of the state-of-the-art deep learning technique is used to make a fusion between these images, which are acquired at multiple b values, and the clinical biomarkers for accurate diagnosis of prostate cancer. The main motivation behind submitting the current manuscript to the "Special Collection on Deep Learning in Medical Imaging" is that the application of deep learning techniques in prostate cancer diagnosis is one of the main areas of this special issue, which is the case of the current submitted manuscript.

Methodology

The basic processing steps of the presented system are summarized in Figure 1. The CAD system begins with delineating the prostate region using a level set–based model. In this model, the evolution of the level set is guided by a nondeterministic speed function that employs nonnegative matrix factorization (NMF). The NMF fuses the DW-MRI intensity information, the probabilistic shape prior, and the spatial voxels interactions. The resulting segmentation accuracy of the developed segmentation model in terms of Dice similarity coefficient and average Hausdorff distance is 86.89% and 5.72 mm, respectively. More information about that segmentation model and comparisons with other segmentation models can be found in our previous work.[17] Then, DW-MRI intensity-based features, such as ADCs are extracted, normalized, smoothed using a generalized Gaussian Markov random field (GGMRF) model, and globally described using the cumulative distribution function (CDF). Those DW-MRI features are integrated with the PSA screening results for better accuracy of diagnosis. Finally, both the PSA-based probabilities and CDFs of the estimated ADCs are fed into a stacked nonnegatively constrained sparse autoencoder (SNCSAE) to predict the diagnosis of the input prostate volume as either benign or malignant through a 2-stage classification.

Imaging Features and Clinical Biomarkers

Key discriminating features are estimated from the delineated prostate region to differentiate between cancerous and benign prostates, as shown in Figure 2. In this work, a DW-MRI intensity-based feature, ADC, is calculated by measuring the difference between 2 DW-MRI data images: one is used as the baseline (b_0) and the other is acquired at a higher b value. The ADC map is the set of ADC values at every voxel and is calculated using the following equation:

$$\mathrm{ADC}(x,y,z) = \frac{\ln \frac{S_0(x,y,z)}{S_1(x,y,z)}}{b_1 - b_0}, \quad (1)$$

where S_0 and S_1 are the intensities obtained, respectively, at the b_0 and b_1 values. It has been demonstrated that the ADC maps are effective in distinguishing between cancerous and benign cases, as benign prostates have a higher average ADC than cancerous ones.[18] The whole ADC maps for all cases at a given b value are then normalized and refined using a GGMRF model

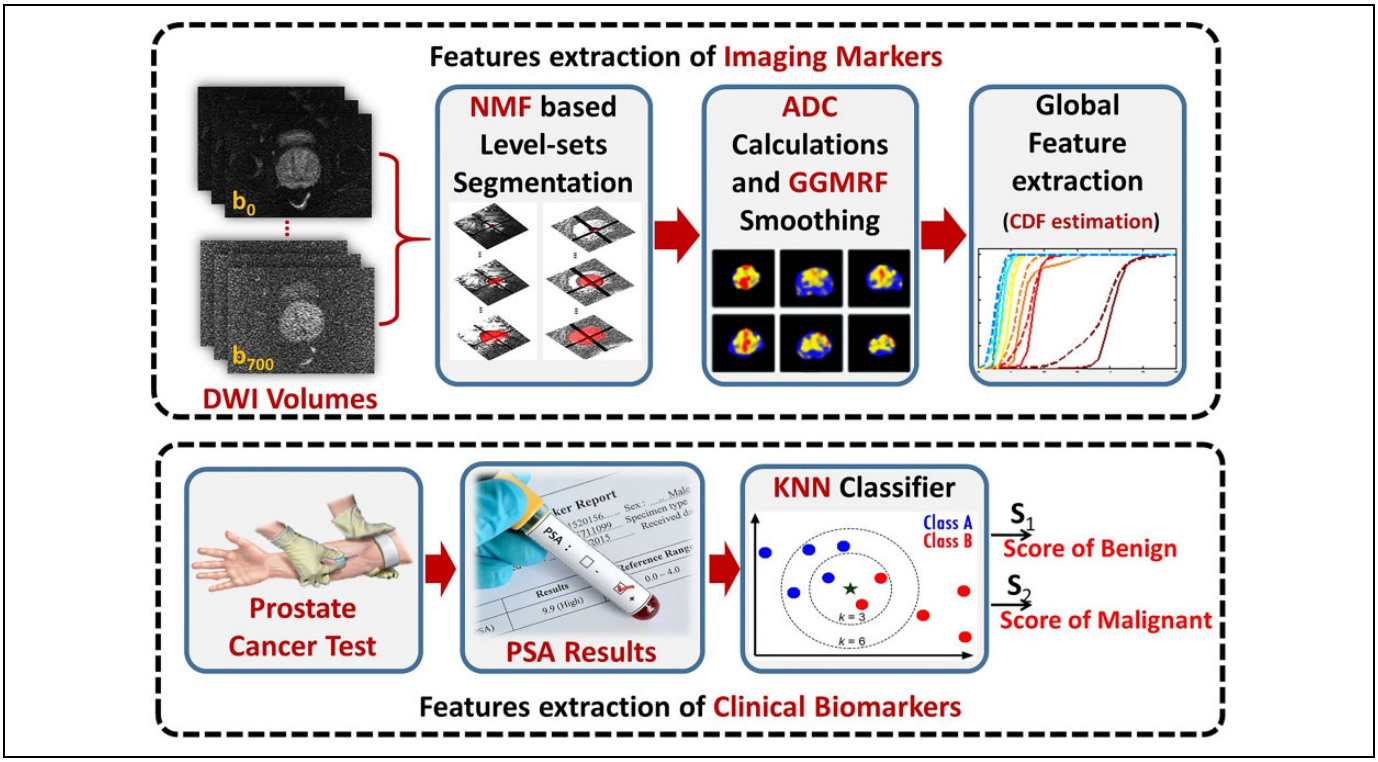

Figure 2. Schematic diagram of both the imaging features and the biomarkers used in the presented system.

with a 26-voxel vicinity to eliminate any discrepancy and maintain continuity. To globally describe the entire ADC volume, the CDFs of the processed ADC maps for each case are calculated. Like ADC maps, these constructed CDFs have the same characteristic of being able to distinguish between cancerous and benign prostates. There are 2 main advantages of using these constructed CDFs as inputs to the SNCSAE-based classifiers instead of the prostate volumes. First, these constructed CDFs have a unified size. Therefore, their use overcomes the challenge related to the variable sizes of different prostate volumes. Second, due to the small size of these CDFs, they speed up the time required for training the SNCSAE-based classifiers as well as the time required for classification. In parallel, the PSA screening results for each case is transformed into a diagnostic probability using a K-nearest neighbor (KNN) classifier. The KNN classifier is considered a good choice for low-dimensional data, which is the case of the PSA screening results. Subsequently, the initial diagnostic probabilities at 7 different b values estimated using SNCSAEs that employ ADC-CDFs are integrated with the PSA-based probabilities to increase the diagnostic accuracy of prostate cancer.

Stacked Nonnegativity Constraint Sparse Autoencoders–Based Classification

In the presented CAD system, the classification of prostates into malignant or benign is obtained by integrating the PSA screening results with a 2-phase structure of SNCSAE. In the first phase, 7 SNCSAE-based classifiers, one for each of the 7 b values (100-700 s/mm^2), are employed to determine an initial classification probability of the prostate case. In the second phase, the resulting initial classification probabilities of the 7 first phase classifiers, in addition to the PSA-based classification probabilities, are then concatenated to form an initial classification probability vector. This vector is fed into another SNCSAE-based classifier to determine the final classification of the prostate case. Each SNCSAE compresses the CDFs, of size 100, at a certain b value inputted into it to grasp the most noticeable variations and is constructed by linking the final hidden layer with a softmax classifier. The SNCSAE is the first pretrained one layer at a time using greedy unsupervised pretraining.[19] Then, a supervised fine-tuning of all SNCSAE layers is performed using error backpropagation to minimize the total loss for the given training data. In the subsequent paragraphs, autoencoder (AE), the basic unsupervised feature learning algorithm, is first introduced. Then, NCSAE, which imposes nonnegativity and sparsity constraints for learning robust feature representations is explained. Finally, SNCSAE, the deep network architecture that is constructed by layer-wise stacking of multiple NCSAE is explained.

Autoencoder, the basic learning component of SNCSAE, consists of 3 layers, which are the input layer, the hidden layer, and the output layer. Each layer has a number of nodes, and a node in a given layer is fully connected to all nodes in the successive layer. The objective of AE is to learn a precise compressed representation of input data that could be used at a later stage to reconstruct the input data. In general, AE has 2 steps, which are encoding and decoding. The encoding layers

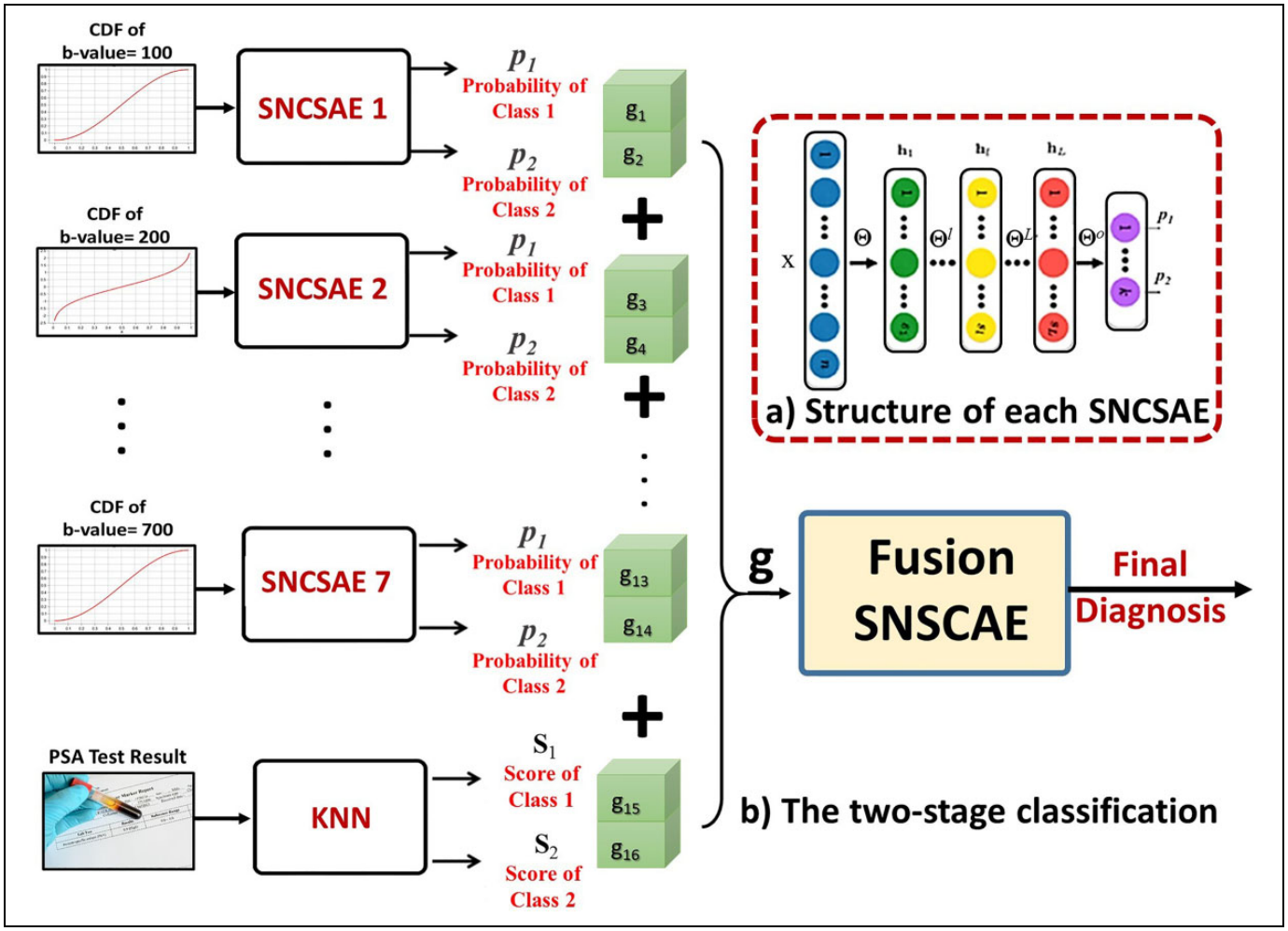

Figure 3. Schematic diagrams of (A) stacked nonnegativity constraint sparse autoencoders (SNCSAE) and (B) the 2-stage classification.

hierarchically decrease the dimension of their inputs into codes to capture the most essential representations, while the decoding layers try to restore the original input from the codes in the hidden layers.

Given an n-dimensional column vector $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ of input data, AE first encodes it into an m-dimensional column vector $\mathbf{h} = [h_1, \ldots, h_m]^{\mathrm{T}}$ of hidden feature representations by the nonlinear activation function $\sigma(\ldots)$ defined as:

$$h_j = \sigma\left(\left(\boldsymbol{\Theta}_j^e\right)^{\mathrm{T}}\mathbf{x}\right) \equiv \sigma\left(\sum_{i=1}^{n}\theta_{j:i}^e x_i\right), \tag{2}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}^e{}_j, \boldsymbol{\Theta}^d{}_i; j = 1, \ldots, s; i = 1, \ldots, n\}$ denotes a set of vectors of trainable coefficients for the layers of encoding (e) and decoding (d) of a given AE, as shown in Figure 3A, T denotes vector transposition, and $\sigma(\phi) = 1/(1 + \exp(\phi))$ is a sigmoid function whose outputs are in the interval [0,1]. Then, AE tries to reconstruct an approximation of the original input from the hidden feature representation. To learn a compressed representation that helps find out concealed structures of high-dimensional data, and to avoid trivial solution of the minimization of the loss function of Equation 3, such as identity transformation, it is required that the hidden layer dimension be less than the input dimension, that is, $m \ll n$. Given a training set of K samples, AE is trained to find the optimal coefficients of all connections by minimizing the loss function that describes the discrepancy between each input vector $\mathbf{x}_k$; $k = 1, \ldots, K$, and its reconstruction vector, b (3) $\hat{\mathbf{x}}_{\Theta:\,k}$ over the entire training set:

$$J_{\mathrm{AE}}(\boldsymbol{\Theta}) = \frac{1}{2K}\sum_{k=1}^{K}\|\hat{\mathbf{x}}_{\boldsymbol{\Theta}:k} - \mathbf{x}_k\|^2. \tag{3}$$

Here, the average sum of squares of differences represents the reconstruction error. The minimization of that reconstruction error indicates that the learned features preserve a significant amount of information about the input which is a required criterion of precise representation of the original input.[20]

Inspired by NMF and by the proofs that neural activity in the human brain is sparse.[20] The NCSAE imposes 2 additional constraints on the basic AE, namely, nonnegativity and sparsity constraints. The nonnegativity constraint

enforces the AE to learn additive part–based representation of its input data, while the sparsity constraint enforces the average activation of each hidden unit over the entire training data set to be infinitesimal to improve the probability of linear separability.[21] As suggested by Hosseini-Asl *et al*,[22] imposing the nonnegativity constraint on AE results in more precise data codes during the greedy layer-wise unsupervised training and improved accuracy after the supervised fine-tuning. Mathematically, the loss function of Equation 3 is extended by the addition of 2 penalty terms to lower the number of negative coefficients and compel sparsity of the NCSAE. Those 2 terms are quadratic negative coefficient penalty, $f(\theta_i) = (\min\{0,\theta_i\})^2$; $i = 1, \ldots, n$, and Kullback-Leibler (KL) divergence, $J_{\mathrm{KL}}(\mathbf{h}_{\Theta}{}^{\mathrm{e}};\rho)$, between the hidden codes, $\mathbf{h}_{\Theta}{}^{\mathrm{e}}$, achieved using the encoding coefficients $\boldsymbol{\Theta}^{\mathrm{e}}$ of the training data set, and a small positive constant value, ρ, close to 0 representing the target average activation. The value of ρ is chosen to be small as small ρ leads to complete and nonredundant features be learned[23]:

$$J_{\mathrm{NCSAE}}(\boldsymbol{\Theta}) = J_{\mathrm{AE}}(\boldsymbol{\Theta}) + \alpha\sum_{j=1}^{s}\sum_{i=1}^{n} f(\theta_{j:i}) + \beta J_{\mathrm{KL}}(\mathbf{h}_{\boldsymbol{\Theta}^{\mathrm{e}}};\rho)\,. \tag{4}$$

The parameters $\alpha \geq 0$ and $\beta \geq 0$ control the amount of contributions of the nonnegativity and the sparsity terms to the total loss function, $J_{\mathrm{NCSAE}}(\boldsymbol{\Theta})$, and

$$J_{\mathrm{KL}}(\mathbf{h}_{\boldsymbol{\Theta}^{e}},\rho) = \sum_{j=1}^{s} h_{\boldsymbol{\Theta}^{e}:j}\log\left(\frac{h_{\boldsymbol{\Theta}^{e}:j}}{\rho}\right) + (1 - h_{\boldsymbol{\Theta}^{e}:j})\log\left(\frac{1 - h_{\boldsymbol{\Theta}^{e}:j}}{1-\rho}\right). \tag{5}$$

Recent studies have demonstrated that a deep architecture has the capability of learning complex and highly nonlinear features from data.[23,24] In order to learn high-level features from data, NCSAE is used as a building block to construct a multilayer architecture of NCSAEs (Figure 3A). In this architecture, the output vector from a low-level NCSAE is used as input to a high-level NCSAE. In addition, the output of the final NCSAE is inputted to a softmax regression classifier. A good technique to train such deep architectures, that does not have the limitations associated with full supervised training, is to first pretrain the network one layer at a time using the unsupervised greedy algorithm. In our system, the first and second NCSAEs, which are the first and second layers of SNCSAE, are pretrained separately to minimize the total loss function of Equation 4. This results not only in decreasing the reconstruction error but also in increasing the number of nonnegative coefficients and the sparsity of the hidden representations. The outputs of the second NCSAE, $\mathbf{h}^{[2]} = \sigma({\boldsymbol{\Theta}^{e}_{[2]}}^{\mathrm{T}}\mathbf{h}^{[1]})$, are inputted to the softmax classifier, as demonstrated in Figure 3A, to estimate the classification of the prostate case at a certain b value as a probability of each output class, $c = 1,2$:

$$p(c;\boldsymbol{\Theta}_{\circ:c}) = \frac{exp\left(\boldsymbol{\Theta}_{\circ:c}^{\mathrm{T}}\mathbf{h}^{[2]}\right)}{exp\left(\boldsymbol{\Theta}_{\circ:1}^{\mathrm{T}}\mathbf{h}^{[2]}\right) + exp\left(\boldsymbol{\Theta}_{\circ:2}^{\mathrm{T}}\mathbf{h}^{[2]}\right)};$$
$$c = 1,2;\ \sum_{c=1}^{2} p\left(c;\boldsymbol{\Theta}_{\circ:c};\mathbf{h}^{[2]}\right) = 1.$$

This unsupervised layer-wise training aims to minimize the negative log-likelihood $J_0(\boldsymbol{\Theta}_0)$ of the training classes, appended with the penalization of negative coefficients:

$$J_{\circ}(\boldsymbol{\Theta}^{o}) = -\frac{1}{K}\sum_{k=1}^{K}\log p(c_k;\boldsymbol{\Theta}_{\circ:c}) + \nu\sum_{c=1}^{2}\sum_{j=1}^{s_2}\theta_{\circ:c:j}. \tag{6}$$

Then, a supervised fine-tuning of all SNCSAE layers follows the unsupervised pretraining to ensure that the learned representations are discriminative.[25] This supervised fine-tuning is performed on the training data using error backpropagation through the layers and with the penalization of the negative coefficients of the softmax classifier only. The parameters $\alpha = 0.003$, $\beta = 5$, and $\rho = 0.5$ were selected empirically based on comparative experiments.

In the first phase of the 2-phase structure of SNCSAE, the initial input data to each of the 7 SNCSAE is composed of the CDFs of size 100 at a certain b value (100-700 s/mm^2). The size of the input vector is decreased by the first layer of SNCSAE to $s_1 = 50$, that is, decreased by the following layer of SNCSAE to $s_2 = 5$, that is, reduced by the final softmax classifier to $s_0 = 2$ probabilities.

In the second phase of the classification, both the PSA-based probabilities and the output probabilities of each of the 7 SNCSAEs of the first stage are concatenated to form an initial probability vector $\mathbf{g} = [g_1, \ldots, g_{16}]$, as sketched in Figure 3B. This probability vector ($\mathbf{g}$) is used as an input to a new SNCSAE to determine the ultimate classification of the prostate case as a probability for each output class, c, using the following formula:

$$p_t(c;\boldsymbol{\Theta}^{t}_{\circ:c}) = \frac{exp\left((\boldsymbol{\Theta}^{t}_{\circ:c})^{\mathrm{T}}g_t\right)}{\sum_{c=1}^{C} exp\left((\boldsymbol{\Theta}^{t}_{\circ:c})^{\mathrm{T}}g_t\right)};\ c = 1,2. \tag{7}$$

Experimental Results

Analyses were conducted on DW-MRI data sets acquired from 18 patients (9 benign and 9 malignant). These patients were diagnosed using biopsy. Biopsy was carried out using a systematic approach with 11 cores taken from the whole prostate. The PSA blood samples were extracted from all 18 patients, 1 week before the participants were scanned by a DW-MRI scanner, using the conventional venous blood draw procedure to get 3 mL from each participant. Figure 4 shows DW-MR images from 2 participants at different b values.

To highlight the benefit of combining clinical biomarkers with DW-MRI features, several experiments that used clinical

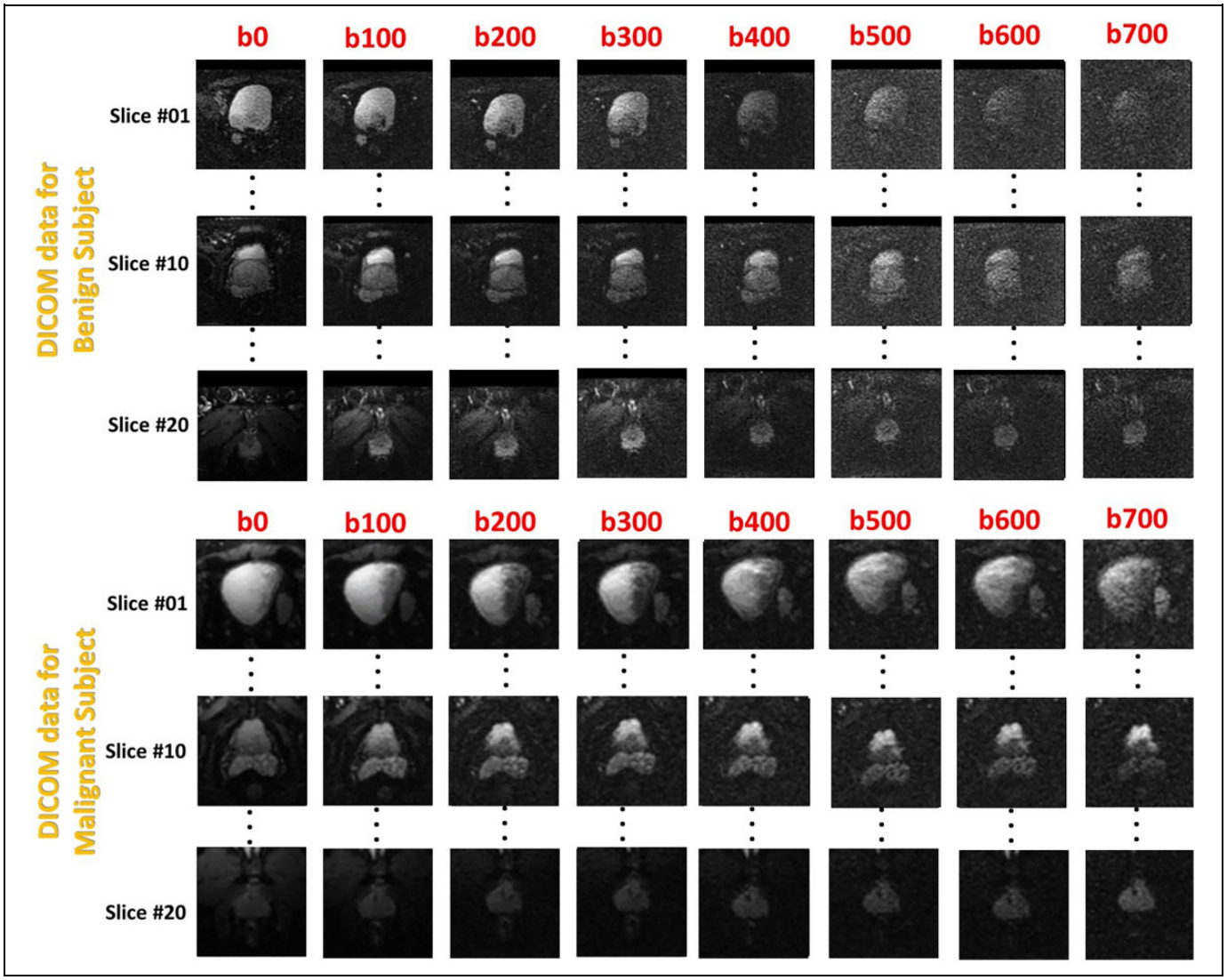

Figure 4. An illustration of diffusion-weighted magnetic resonance imaging (DW-MRI) digital imaging and communications in medicine (DICOM) images for 2 different cases (1 benign and 1 malignant) at different b values.

biomarkers only, DW-MRI features only, or a combination of both clinical biomarkers and DW-MRI features, have been conducted. For the DW-MRI experiment, the discriminating characteristics of benign and cancerous cases are captured from the DW-MRI data sets by training 7 different SNCSAEs, one SNCSAE for each of the 7 different b values (100, 200, . . . , 700 s/mm^2). The features adopted for the diagnosis are the CDFs of the processed ADC volumes, of the delineated prostates, estimated at 7 different b values, as discussed previously. A combination of lower b values and higher b values is utilized because lower b values show perfusion while higher b values show diffusion. According to the literature, both perfusion and diffusion can be used to discriminate malignant tumors from benign ones. The middle in-between b values are mix, that is, they do not represent clear perfusion and do not represent clear diffusion. This highlights why a combination of lower and higher b values shows good results for early diagnosis of prostate cancer. This is why we start to acquire data at multiple b values to capture both perfusion and diffusion to lead to more accurate results. An illustration of the ADC color maps of 2 participants at different b values is illustrated in Figure 5. In order to test the effect of each feature (CDFs of each b value), each SNCSAE of the first-stage of the classification is first trained and tested for each feature individually. To evaluate the accuracy of this experiment, a leave-one-subject-out (LOSO) cross-validation is performed for each SNCSAE with all 18 DW-MRI data sets. The diagnostic accuracy for each SNCSAE is reported in Table 1. In order to get a global classification decision based only on DW-MRI, all diffusion-based probabilities from the 7 SNCSAEs are concatenated to form an input vector that is fed into a new SNCSAE. The resulting accuracy after this fusion is 88.89% (sensitivity = 88.89% and specificity = 88.89%). Moreover, a 3-fold cross-validation is performed using the DW-MRI data sets at each b value. The diagnostic accuracy at each b value is reported in Table 2.

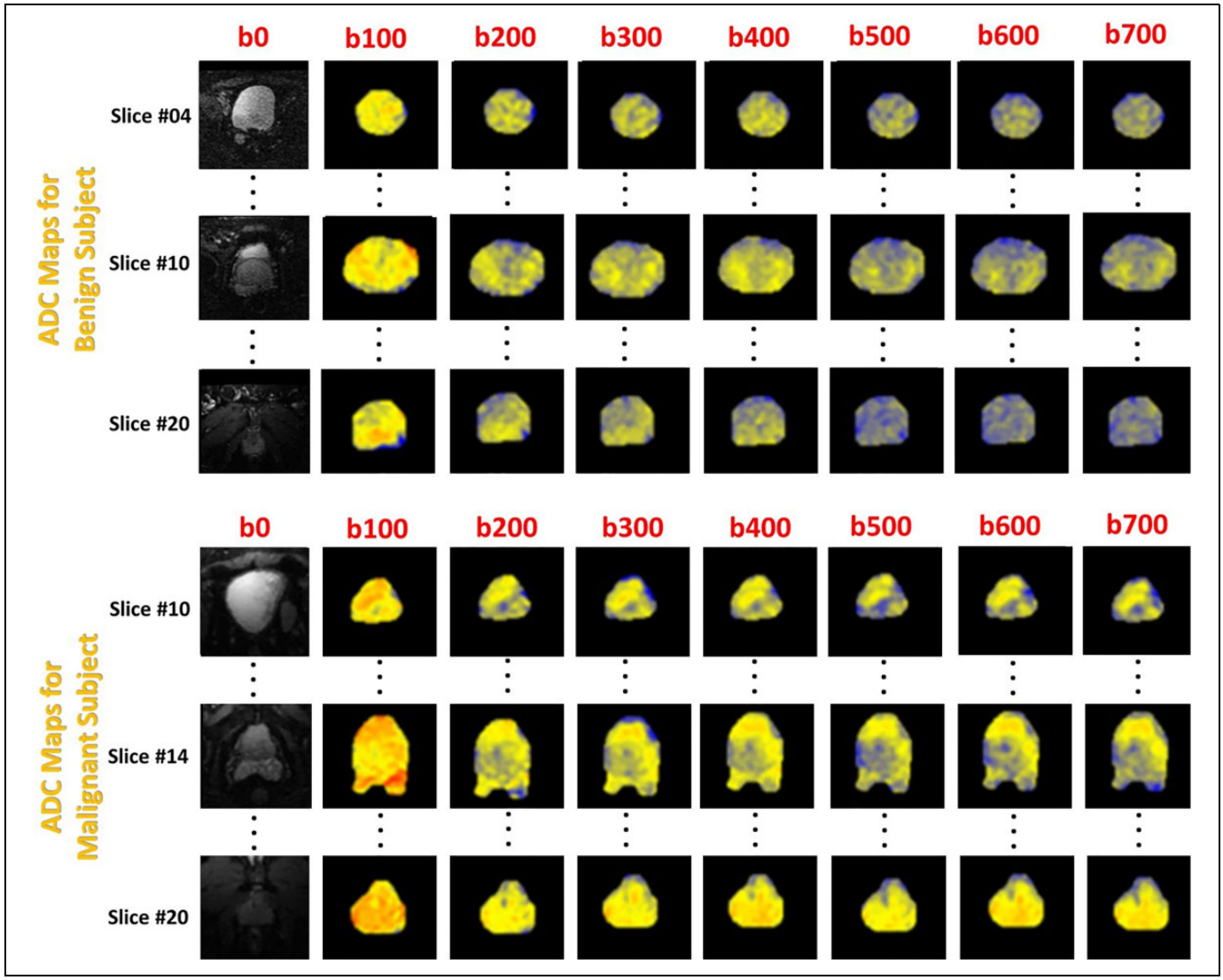

Figure 5. An illustration of the corresponding apparent diffusion coefficient (ADC) color maps for 2 cases (1 benign and 1 malignant) at different *b* values.

Table 1. Performance Results of SNCSAE-Based Classifiers at the 7 *b* Values Using a LOSO Cross-Validation.

SNCSAE	Accuracy (%)	Sensitivity (%)	Specificity (%)
SNCSAE 1 (*b* value = 100)	77.8	77.8	77.8
SNCSAE 2 (*b* value = 200)	66.6	77.8	55.6
SNCSAE 3 (*b* value = 300)	72.2	77.8	66.7
SNCSAE 4 (*b* value = 400)	72.2	77.8	66.7
SNCSAE 5 (*b* value = 500)	72.2	77.8	66.7
SNCSAE 6 (*b* value = 600)	83.3	88.9	77.8
SNCSAE 7 (*b* value = 700)	83.3	88.9	77.8

Abbreviations: LOSO, leave-one-subject-out; SNCSAE, stacked nonnegativity constraint sparse autoencoders.

Table 2. Diagnostic Accuracy Using 3-Fold Cross-Validation at the 7 *b* Values.

SNCSAE	First Fold (%)	Second Fold (%)	Third Fold (%)	Average (%)
SNCSAE 1	66.7	83.3	66.7	72.2
SNCSAE 2	66.7	83.3	50	66.7
SNCSAE 3	66.7	50	83.3	66.7
SNCSAE 4	66.7	50	83.3	66.7
SNCSAE 5	83.3	50	66.7	66.7
SNCSAE 6	66.7	50	100	72.2
SNCSAE 7	66.7	100	66.7	77.8

Abbreviation: SNCSAE, stacked nonnegativity constraint sparse autoencoders.

For the PSA screening experiment, the PSA screening result of each case is transformed into a diagnostic probability using a KNN-based classifier. To apply a KNN classifier that identifies the prostate status, again we use an LOSO cross-validation for each participant. The resulting accuracy reported in Table 3 emphasizes the need to combine other features with the PSA screening results to improve the system accuracy.

Finally, the results from both the DW-MRI and PSA experiments are fed into the fusion SNCSAE to determine the final diagnosis of the input prostate as benign or malignant through a 2-stage classification network. The overall classification accuracy, sensitivity, and specificity are 94.4%, 88.9%, and 100%, respectively, for the LOSO cross-validation. The overall classification accuracy, sensitivity, and specificity are 88.9%, 77.8%, and 100%, respectively, for the 3-fold cross-validation. These results emphasize the advantage of integrating DW-MRI with PSA results for prostate cancer diagnosis. The LOSO cross-validation after fusing the clinical biomarkers with the imaging makers at the different b values resulted in one misclassified case. Although the 3-fold cross-validation resulted in one more misclassified case. These 2 misclassified cases will need more investigation. Their PSA values are in the range of benign cases whereas their biopsy shows that they are malignant. We asked the physician to pay more attention to these cases, especially the case that is misclassified by both the LOSO and the 3-fold. After radical prostatectomy, more investigation can be done on the histology level. Recently, many researchers began to utilize convolutional neural network (CNN) for prostate cancer detection.[26-30] To overcome the limitation of the dependence of the diagnostic accuracy on the segmentation accuracy, our next step will be feeding the imaging markers into a CNN instead of SNCSAE.

Table 3. Performance Results of the KNN Classifier Using PSA Screening Results.

Classifier	Accuracy	Sensitivity	Specificity
KNN	77.78%	55.56%	100%

Abbreviations: KNN, K-nearest neighbor; PSA, prostate-specific antigen.

Table 4. Performance Results of the Presented CAD System and 2 Classifiers (RF and RT).

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
SNCSAE	94.4	88.9	100	0.98
RF	88.9	88.9	88.9	0.97
RT	88.9	100	77.8	0.88

Abbreviations: AUC, area under the curve; CAD, computer-aided diagnostic; RF, random forest; RT, random tree; SNCSAE, stacked nonnegativity constraint sparse autoencoders.

To manifest the advantage of using an SNCSAE-based classifier, we have also conducted a comparison between the presented classifier and 2 state-of-the-art classifiers, RF and random tree (RT).[31] The resulting accuracy, sensitivity, specificity, and AUC of these classifiers are provided in Table 4. According to the results of that table, the performance of the SNCSAE-based classifier is better than the performance of both the RF and the RT, indicating the promising diagnostic capabilities of the presented CAD system. The corresponding ROC curve of the presented SNCSAE-based classifier and the other 2 classifiers are illustrated in Figure 6. The 95%

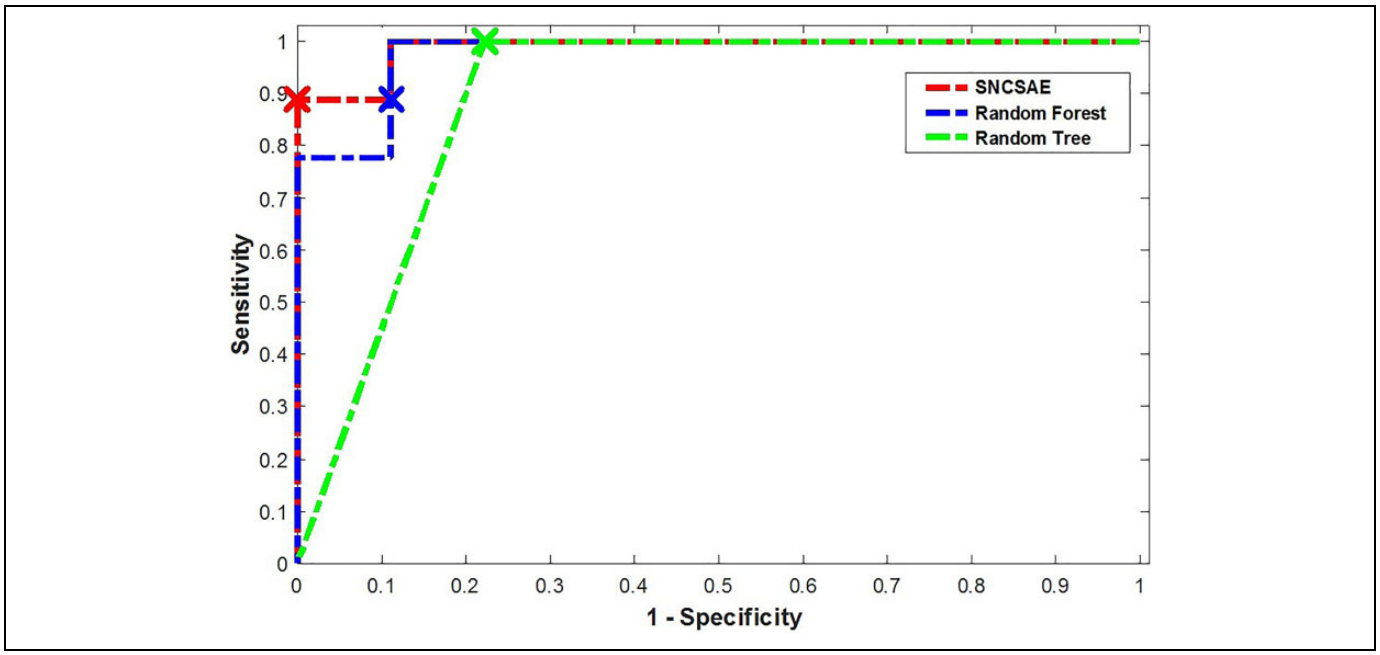

Figure 6. The receiver operating characteristic (ROC) curve of the presented stacked nonnegativity constraint sparse autoencoders (SNCSAE)-based classifier, random forest (RF), and random tree (RT).

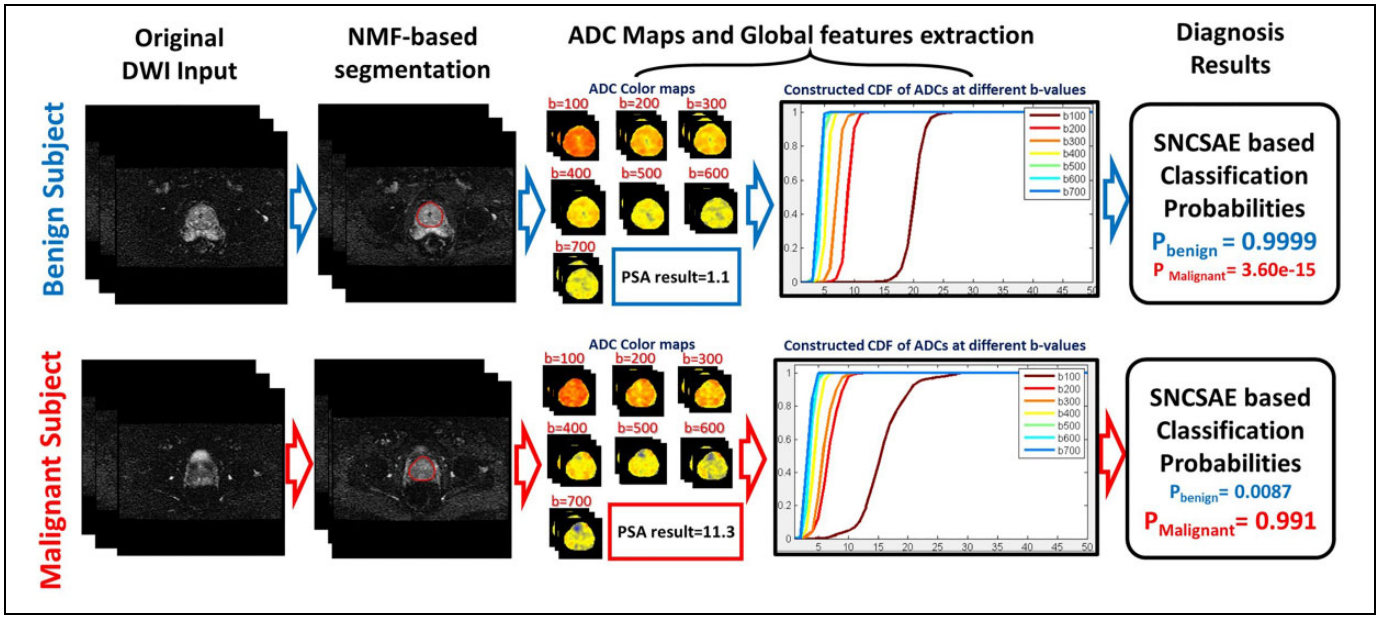

Figure 7. A step-by-step illustration of the presented framework for 2 different cases (1 benign and 1 malignant).

confidence interval (CI) is computed using the bootstrapping technique.[32] A random sample of 18 participants is selected with replacement and the corresponding AUC is computed. This operation is repeated 100 times. The effect of the replacement aspect on the results is that if the misclassified participants are selected more than 1 time, then the resulting performance in terms of the AUC will be reduced. The resulting 95% CI ranges from 0.79 to 1. In our case, the number of the used participants is limited. However, the upper bound of 95% CI equals one, which indicates that the accuracy of the presented method, in terms of sensitivity and specificity, can reach 100%, in case of increasing the number of participants. It is important to mention that the presented system has been tested using all currently available data with our group. The number of participants of the used data set is limited. The focus of our future work will be on collecting more participants to validate our system with a larger data set to highlight the pros and cons of the presented approach. Figure 7 shows the pipeline processing steps for 2 different prostate cases (1 benign and 1 malignant).

Conclusions

In summary, this article presents a noninvasive CAD system that integrates PSA screening results in addition to DW-MRI-based features in order to diagnose prostate cancer. The presented CAD system uses a hybrid approach to effectively integrate level sets and NMF for prostate segmentation, constructs global features using the ADC-CDFs of the delineated prostate regions to differentiate between benign and malignant cases, and utilizes a 2-phase SNCSAEs on both the ADC-CDFs and PSA-based classification probabilities to create a more robust diagnosis platform for prostate cancer. The presented CAD system successfully yields an accurate (94.4%) diagnosis platform using a DW-MRI data set collected at 7 different b values (100, 200, . . . , 700 s/mm^2) from 18 participants. The focus of the future work will be on augmenting the number of participants used in both the learning and the evaluation of the presented CAD system to enhance and confirm the robustness of the results obtained in this study. Another future work is to acquire DW-MRI data sets at higher b values in order to examine their effect on the accuracy of the system.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66(1):7-30.
2. Ferlay J, Shin HR, Bray F, et al. *GLOBOCAN 2008: Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10.* Lyon, France: International Agency for Research on Cancer; 2010.
3. Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. *J Am Board Fam Pract*. 2003;16(2):95-101.
4. Dijkstra S, Mulders PF, Schalken JA. Clinical use of novel urine and blood based prostate cancer biomarkers: a review. *Clin Biochem* 2014;47(10-11):889-896.
5. Davis M, Sofer M, Kim SS, Soloway MS. The procedure of transrectal ultrasound guided biopsy of the prostate: a survey of patient preparation and biopsy technique. *J Urol*. 2002;167(2): 566-570.

6. Hricak H, Choyke PL, Eberhardt SC, Leibel SA, Scardino PT. Imaging prostate cancer: a multidisciplinary perspective 1. *Radiology*. 2007;243(1):28-53.
7. Applewhite JC, Matlaga B, McCullough D, Hall MC. Transrectal ultrasound and biopsy in the early diagnosis of prostate cancer. *Cancer Control*. 2000;8(2):141-150.
8. Tan CH, Wang J, Kundra V. Diffusion weighted imaging in prostate cancer. *Eur Radiol* 2011;21(3):593-603.
9. Reda I, Shalaby A, Khalifa F, et al. Computer-aided diagnostic tool for early detection of prostate cancer. In: *IEEE International Conference on Image Processing*. Phoenix, AZ: IEEE, September 25-28, 2016, pp. 2668-2672.
10. Tamada T, Sone T, Jo Y, Yamamoto A, Ito K. Diffusion-weighted MRI and its role in prostate cancer. *NMR Biomed*. 2014;27(1):25-38.
11. 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.
12. Hambrock T, Vos PC, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging effect on observer performance. *Radiology*. 2013;266(2):521-530.
13. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging*. 2014;33(5):1083-1092.
14. Kwak JT, Xu S, Wood BJ, et al. Automated prostate cancer detection using T2-weighted and high-b-value diffusion-weighted magnetic resonance imaging. *Med Phys*. 2015;42(5):2368-2378.
15. Peng Y, Jiang Y, Yang C, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology*. 2013: 267(3):787-796.
16. Boesen L, Chabanova E, Løgager V, Balslev I, Thomsen HS. Apparent diffusion coefficient ratio correlates significantly with prostate cancer gleason score at final pathology. *J Magn Reson Imaging*. 2015;42(2):446-453.
17. McClure P, Khalifa F, Soliman A, et al. A novel NMF guided level-set for DWI prostate segmentation. *J Comput Sci Syst Biol*. 2014;7(6):209-216.
18. Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. *Radiology*. 2013;268(2):318-322.
19. Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks. In: *Advances in Neural Information Processing Systems*. Vancouver, BC, Canada: MIT Press; 2006: 153-160.
20. Han J, Zhang D, Wen S, Guo L, Liu T, Li X. Two-stage learning to predict human eye fixations via SDAEs. *IEEE Trans Cybern*. 2016;46(2):487-498.
21. Boureau Y-L, Cun YL, et al. Sparse feature learning for deep belief networks. In: *Advances in Neural Information Processing Systems*. Vancouver, BC, Canada: MIT Press; December 3-6, 2007, pp. 1185-1192.
22. Hosseini-Asl E, Zurada JM, Nasraoui O. Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints. *IEEE Trans Neural Networks Learn Syst*. 2016;27(12):2486-2498.
23. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(8):1798-1828.
24. Yan K, Li C, Wang X, et al. Comprehensive autoencoder for prostate recognition on MR images. In: *IEEE International Symposium on Biomedical Imaging*. Prague, Czech Republic: IEEE, April 13-16, 2016, pp. 1190-1194.
25. Rota Bulo S, Kontschieder P. Neural decision forests for semantic image labelling. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH: IEEE; June 24-27, 2014, pp. 81-88.
26. Tsehay YK, Lay NS, Roth HR, et al. Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. In: *SPIE Medical Imaging*. Orlando, FL, February 11-16, 2017, pp. 1013405-1013405.
27. Le MH, Chen J, Wang L, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. *Phys Med Biol*. 2017:62(16): 6497-6514.
28. Clark T, Wong A, Haider MA, et al. Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted MR images. In: *International Conference Image Analysis and Recognition*. Montreal, Canada: Springer, July 5-7, 2017, pp. 97-104.
29. Chung AG, Shafiee MJ, Kumar D, et al. Discovery radiomics for multi-parametric MRI prostate cancer detection. *arXiv preprint*. 2015; arXiv:1509.00111.
30. Liu S, Zheng H, Feng Y, et al. Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. *arXiv preprint* 2017; arXiv:1703.04078.
31. Hall M, et al. The WEKA data mining software: an update. *SIGKDD Explor Newsl*. 2009;11(1):10-18.
32. Skalska H, Freylich V. Web-bootstrap estimate of area under ROC curve. *Austrian J Stat*. 2016;35(2-3):325-330.