Prostate Cancer Classification with Multi-parametric MRI Transfer Learning Model

Yixuan Yuan*

Department of Electronic Engineering,

City University of Hong Kong, Hong Kong and

Department of Radiation Oncology, Stanford University, Stanford 94305, USA

Wenjian Qin†

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences,

Shenzhen 518055, People's Republic of China

Mark Buyyounouski, Bulat Ibragimov, Steve Hancock, Bin Han and Lei Xing‡

Department of Radiation Oncology, Stanford University, Stanford 94305, USA

Bin Han and Lei Xing are co-corresponding authors.

(Email:hanbin@stanford.edu, lei@stanford.edu)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.13367

Abstract

Purpose: Prostate cancer classification has significantly impact on the prognosis and treatment planning of patients. Currently, the classifying is based on the Gleason score analysis of biopsied tissues, which is neither accurate nor risk-free. This study aims to learn discriminative features for prostate images and assist physicians to classify prostate cancer automatically.

Methods: We develop a novel multi-parametric magnetic resonance transfer learning (MPTL) method to automatically stage prostate cancer. We first establish a deep convolutional neural network with three branch architectures, which transfer pre-trained model to compute features from multi-parametric MRI images (mp-MRI) : T2w transaxial, T2w sagittal and apparent diffusion coefficient (ADC). The learned features are concatenated to represent information of mp-MRI sequences. A new image similarity constraint is then proposed to enable the distribution of the features within the same category in a narrow angle region. With the joint constraints of softmax loss and image similarity loss in the fine-tuning process, the MPTL can provide descriptive features with intraclass compactness and interclass separability.

Results: Two cohorts: 132 cases from our institutional review board approved patient database and 112 cases from the PROSTATEx-2 Challenge are utilized to evaluate the robustness and effectiveness of the proposed MPTL model. Our model achieved high accuracy of prostate cancer classification (accuracy of 86.92%). Moreover, the comparison results demonstrate that our method outperforms both hand-crafted feature based methods and existing deep learning models in prostate cancer classification with higher accuracy.

Conclusion: The experiment results showed that the proposed method can learn discriminative features for prostate images and classify the cancer accurately. Our MPTL model could be further applied in the clinical practice to provide valuable information for cancer treatment and precision medicine.

Keywords: Multi-parametric magnetic resonance transfer learning (MPTL), prostate cancer classification, image similarity constraint, Gleason score.

* yxyuan.ee@cityu.edu.hk

† qinwenjian@gmail.com

‡ mbuyyou@stanford.edu, bulat@stanford.edu, shancock@stanford.edu, hanbin@stanford.edu, lei@stanford.edu

I. INTRODUCTION

Prostate cancer is the second most common cancer in men worldwide. According to statistics of American Cancer Society, 164,690 new cases of prostate cancer will be examined and 29,430 cases of cancer death will occur in the United States in 2018 [1]. Accurate grading and classifying of prostate cancer are critical for the selection of the most suitable treatment, and ultimately reducing prostate cancer morbidity and mortality.

Standard technique for classifying of prostate cancer is based on the Gleason score, identified by biopsy samples acquired under transrectal ultrasound guidance. However, the sensitivity of the standard biopsy for detection of prostate cancer, has been reported to be as low as 40% [2]. Moreover, the Gleason classification system suffers from inter- and intra-observer variability due to its heavy reliance on human interpretation. Recent advances in prostate multi-parametric magnetic resonance imaging (mp-MRI) have improved cancer grading and classifying [3]. Briefly, high-grade prostate cancer tends to have denser glandular structures and a prominent desmoplastic stromal reaction, which can be better detected by machine learning process for MRI images. This work is aimed at designing an automated calculation method for prostate cancer classification, which can be used as objective references to pathologists.

Many machine learning approaches [4–16] have been proposed for prostate cancer classification with MRI images. Most of these methods used low-level radiomics features based on the previous clinical experiences, which could not take full advantage of the complete information included in MRI images [17]. Moreover, the features utilized in previous methods were often calculated with unsupervised strategies, which may include useless and redundant information or even ignore some important clues. Recently, deep learning methods attracted great attention and achieved inspiring performance in recognition and classification tasks of both natural images [18], and biomedical applications [19–21]. They can jointly learn features and train classifiers. Inspired by the excellent performance and enormous potential of emerging deep learning techniques, here we utilize the deep learning for classifying prostate cancer.

There are a few challenges to directly apply existing deep learning models from the domain of natural color images to medical images. First, the number of collected prostate MRI images is very limited due to the privacy issue. Additionally annotation must be performed by multiple specialists to ensure its validity. The lack of data makes the training very difficult and may lead to overfitting. In order to obtain a good performance with limited training samples, the transfer

learning strategy [22, 23], which applying the well trained deep models from other image datasets, should be utilized. Secondly, different MRI sequences reveal different aspects of prostate cancer, and an effective way of integrating independent sets of information must be in place. For example, the apparent diffusion coefficient (ADC) values have negative correlation with the Gleason score of prostate carcinoma [4]. T2-weighted (T2w) sequence is optimal for depicting the zonal anatomy of the prostate and can be used for detailed assessment of the prostatic fossa and seminal vesicles [24]. A complete characterization of prostate cancers is only achievable by aggregating the information extracted from mp-MRI images. Thirdly, the traditional transfer learning model considers each image independently without including the intra-category correlation information. Thus, the learned deep features do not realistically reflect the clinical decision-making process [25]. But in the real case, images in the same category should possess close deep features. Incorporation of image similarity constraint is highly desirable to learn discriminative deep features.

In this paper, we develop a novel multi-parametric magnetic resonance transfer learning (MPTL) method to automatically stratify prostate cancer. The proposed model is composed of three branches of transfer learning to extract features from mp-MRI sequences: T2w transaxial, T2w sagittal and ADC. The learned features from these categories are aggregated in the model to characterize the MRI images. In the final fine-tuning stage, we introduce a novel image similarity loss function to enable the learned features within the same category to be close. With inclusion of this constraint and softmax loss constraint, the learned joint-features preserve the discriminative representation and enlarge the feature margin between categories. Therefore, the discriminative ability of learned features is highly improved. Our main contributions are summarized as follows.

1. We propose a novel transfer learning method to exploit mp-MRI sequences to stratify prostate cancer. Compared with radiomics features, the joint learned features are more discriminative and improve the robustness of prostate cancer classification.

2. Instead of using a single MRI sequence to characterize prostate cancer region, the mp-MRI sequences are jointly considered in our end-to-end MPTL model. Thus the proposed method can encode richer image information and make decision synergistically by integrating all available information.

3. We embed a novel image similarity constraint into the transfer learning model, which enables small intra-variances among image features. Thus the extracted high-level features

preserve image category information, and enhance the performance of prostate cancer classification. To the best of our knowledge, this work represents the first attempt to introduce image similarity information to deal with prostate cancer classification, which are crucial in the practical clinical applications.

4. We have done extensive experiments on the prostate cancer datasets to show the success of the proposed MPTL model. The role of each component in the model is examined carefully.

II. RELATED WORK

A. Existing Prostate Cancer Diagnosis Methods

Many methods have been proposed to classify prostate cancer. Most of them utilized conventional machine-learning techniques, which typically extract low-level radiomics features to characterize images and train a separate classifier [4–16]. Fehr et al. [8] utilized first-order (mean, SD, skewness, and kurtosis) and second-order texture features (Haralick Features from gray level co-occurrence matrix (GLCM)) to assess the Gleason scores. Vignati et al. [9] used the contrast and homogeneity of GLCM texture features on T2w images and ADC maps to help differentiating between patients with different Gleason scores. Sobecki et al. [11] applied the contrast, homogeneity, energy, angular second moment and correlation features from GLCM to present features of prostate cancers. Then a multi-layer feed-forward Artificial neural network (ANN) with stochastic gradient decent was implemented to classify prostate lesions. The method [5] proposed Probabilistic Support Vector Machine (SVM) model for prostate cancer diagnosis based on mp-MRI images. This method evaluated the class labels through minimizing a hinge loss cost function as well as insensitive cost function together with a minimum norm constraint. A thorough review of radiomics features was recently performed [7].

Recently, deep learning achieved inspiring performance in image classification with discriminative and representative features. While lots of works have been proposed to analyzing natural images, only a handful of works deal with the problem of prostate cancer classification with deep learning methods [26–31]. Reda et al. [26] trained a stacked auto-encoder network with non-negativity constraint algorithm with a logistic regression classifier to distinguish the prostate tumor as benign or malignant with ADC images. Kiraly et al. [28] proposed multi-channel image-to-image convolutional encoder-decoders to localize lesions and then output different tumor classes.

Chen et al. [29] used transfer learning approach with ImageNet pre-trained inceptionV3 and Vgg-16 model to classify prostate cancers from mpMRI scans. Le et al. [30] proposed a new similarity loss function in the multimodal convolutional neural networks for prostate cancer diagnosis to enforce the features extracted from ADC and T2W images consistent. However, these methods [26–31] utilized existing deep learning models to conduct prostate cancer classification task. The specific problems (such as limited data) associated with the medical images were tacitly ignored, which may lead to failing or overfitting in the training procedure.

B. Transfer Learning

In medical imaging, the labeled datasets are small compared to natural image datasets, making the direct training of a convolutional neural network difficult. The transfer learning [22, 23] is an alternative way to deal with limited data. The knowledge gained from large datasets in natural images are transferred to solve the medical imaging problems. The transfer learning methodologies can be classified into two distinctive subdivisions.

1) Feature extractor: This strategy utilizes a pretrained model as a feature extractor. It removes the last full connected layer (classifier layer), and then treats the other layers as a fixed feature extractor to adapt to a new task. This strategy only trains a new classifier instead of the whole network, which significantly speeds up the training process. However, this strategy separates the training process of the feature extraction and the classifier, which decreases the discriminative performance of classification.

2) Fine-Tuning: The fine-tuning strategy not only replaces and retrains the classifier on the new dataset, but also fine-tune the weights of the pretrained network by continuing the back-propagation.

Transfer learning has been extensively studied over the past few years, especially in the field of medical image analysis [32, 33]. However, research that incorporates transfer learning model to deal with prostate cancer classification is sparse [29, 30]. Moreover, these methods treated each prostate cancer image separately and ignored the image similarity information.

III. MULTI-PARAMETRIC TRANSFER LEARNING FOR PROSTATE CANCER

In this paper, we propose a novel multi-parametric magnetic resonance transfer learning (MPTL) model for automated prostate cancer classifying. The workflow of the proposed model is shown in Fig. 1. We first construct three streams of transfer learning model to learn features from different mp-MRI sequences: T2w transaxial, T2w sagittal and ADC. Then these features are concatenated together to obtain discriminative descriptor for prostate cancer regions. Finally we introduce a novel joint loss function to fine-tune MPTL model, where the softmax loss globally enables the features of different classes to stay separate, and the image similarity loss effectively compacts the learned deep features within the same category into a narrow angle region. Hence, the discriminative ability of the learned features could be improved, further enhance the performance of prostate cancer classifying.

A. Single Sequence Transfer Learning to Get Features

Due to the limitation of the size of prostate cancer data, we adopt transfer learning scheme instead of training a whole deep learning neural network. There are many successful deep learning architectures include AlexNet [34], GoogleNet [35], and VGGNet [36]. These models have already been trained on ImageNets, and can be downloaded as the pretained networks. In our paper, AlexNet is chosen. AlexNet consists of five convolutional layers and three fully connected layers. The max-pooling layers are followed with first, second, and fifth convolutional layers to reduce the overfitting. A rectified linear unit nonlinearity is applied to each convolutional and fully connected layer to enable faster training.

By applying the AlexNet on mp-MRI sequences, the features of T2w transaxial, T2w sagittal and ADC images can be calculated as $f^l_{T2w\_t}$, $f^l_{T2w\_s}$, f^l_{ADC}. In this way, we transfer the information learned from ImageNet to characterize prostate cancer images.

B. Feature Fusion with Multi-Sequence MRI

Different MRI sequences reveals different aspects of prostate cancer. It is reported that T2w and ADC provide different and complementary information, and their fusion can effectively improve accuracy of prostate cancer diagnosis [37]. In order to exploit the correlation among different MRI sequences, it is better to learn the MRI information jointly in an end-to-end framework.

In our paper, we feed simultaneously $f^l_{ADC}(s_n)$, $f^l_{T2}(s_n)$, $f^l_{T1}(s_n)$ to the fully connected layers. The output of $(l+1)th$ layer is

$$f^{l+1} = g\{W^{l+1}.[f^l_{T2w\_t} \oplus f^l_{T2w\_s} \oplus f^l_{ADC}] + b^{l+1}\} \tag{1}$$

where W^{l+1} and b^{l+1} are the weight matrix and bias of the fully connected layer, respectively. $\oplus$ means concatenating the features from different sequences. f^{l+1} represents the feature output after five convolutional layers and conveys both T2w and ADC information. Such fusion strategy enables the feature learning process of different modalities mutually affected by each other, yielding consistent and discriminant features related to prostate cancers. After two fully connected layers, the output is regarded as the final joint features.

C. Fine-tuning Training Strategy with Category Similarity Loss Function

After obtaining the joint features from mp-MRI sequences, an additional fine-tuning strategy is used to improve the performance of prostate cancer classifying. It uses the parameters learned from a previously trained network on a specific data set and then adjusts the parameters for our prostate cancer data.

To fine-tune the model, we keep the structure of pretrained model unchanged and remove only the last classifier layer. In the original AlexNet, the last layer computes the probability of 1000 classes of the ImageNet data set. In this paper, we classify the prostate cancer into high grade and low grade cancer. Thus the number of nodes in the last layer of our method is 2. With minor modification of the number of output layers, the strategy can be extended to multiple levels of classification in the future.

Given a training set $T = \{T_1, T_2, ..., T_M\}$, where T_m represents training samples belonging to the mth category. For a training sample x_i belong to category m, we define its corresponding label as l_i with only one element being 1 at the m^{th} position. The traditional classification layer uses the soft-max function as follows to calculate the loss values,

$$L = \frac{1}{N} \sum_{x_i \in T} l_i \log(\hat{l}_i), \tag{2}$$

where $\hat{l}_i$ represents the probability of image x_i being correctly classified as category l_i. N defines the number of training samples. This loss function calculates the differences of the predicted output labels from the true ones.

However, this loss function treats each image equally. In the prostate cancer classification problem, the images from the same category should preserve similar features. Therefore, we define an image similarity constraint and add it to the loss function to describe the feature relationships among categories. The new loss function can be represented as

$$L = \frac{1}{N}\sum_{i=1}^{N} l_i \log(\hat{l_i}) + \frac{1}{N}\sum_{i=1}^{N}(1 - \frac{f_i^T c_m}{\|f_i\|\|c_m\|}), \tag{3}$$

where the first item is the traditional soft-max loss function. The second item represents the image similarity item. f_i is the final feature of image x_i and $c_m = \frac{1}{N_{x_i \in T_m}} \sum_{x_i \in T_m} f_i$ calculates the mean features for the m^{th} category. This loss function minimizes the angle of the feature and its intra-class center and enables the features with the same category close to each other. The joint training of the softmax loss and the image similarity loss reduces the feature variances among same classes effectively, and thus enhances the discriminative performance of the deeply learned features.

During the training process, the stochastic gradient descent (SGD) [38] is used to address the optimization problem of the loss minimization. After fine-tuning the MPTL model, we can conduct the prostate cancer classifying task. Since the feature descriptor and the soft-max classifier with image category information are trained together, the MPTL model could characterize images well and further improve the classification results.

D. Evaluation Criteria

The criteria for the prostate cancer classification is accuracy (ACC), recall (Rec), precision (Pre) and F1-score (F1). They are calculated by:

$$Acc = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \tag{4}$$

$$Rec = \frac{N_{TP}}{N_{TP} + N_{FN}}, \tag{5}$$

$$Pre = \frac{N_{TP}}{N_{TP} + N_{FP}}, \tag{6}$$

$$F1 = \frac{2 \times Rec \times Pre}{Rec + Pre}, \tag{7}$$

where $N_{TP}, N_{TN}, N_{FP}, N_{FN}$ denote the number of true positives, true negatives, false positives, and false negatives. For prostate cancer classification, images with high grade cancer are defined as positives while the ones with low grade cancer are recognized as negatives. Moreover, the receiver

operating characteristic (ROC) curve is plotted for evaluation. The areas under ROC curve (AUC) is also utilized to evaluate the classification performance of prostate cancer.

IV. RESULTS

A. Datasets and Experiment Setup

We applied two cohorts to evaluate the proposed MPTL model. The first one came from PROSTATEx-2 Challenge. 112 Gleason score findings with T2-weighted images (transaxial and sagittal), Ktrans images (computed from dynamic contrast-enhanced images), and ADC images (computed from diffusion-weighted imaging) are collected. The prostate MR scans were acquired in the Prostate MR Reference Center, Radboud University Medical Centre and the detailed data set was previously documented in [6]. The second dataset is collected from Stanford University Medical Center. This dataset consists of 132 prostate cancer cases, obtained under ethical Institutional Review Board (IRB) approval. All patients had a suspicious screening result and a core biopsy yielding Gleason Score (GS). The prostate cancer regions were segmented independently on both the T2w and the ADC-map by experienced radiologists. The ground truth of the Gleason score was confirmed by pathology. We combined these two data sets to conduct experiments. After minor data cleaning, we selected 221 prostate cancer cases. Each case includes a T2w transaxial image, a T2w sagittal image and an ADC map. The annotations include two-class labels: clinically high grade cancer (Gleason score=4+3, 4+4, 3+5, 5+3) and low grade cancer (Gleason score=3+3, 3+4) [4]. To be specific, our dataset consists of 123 cases with high grade cancer and 98 cases with low grade cancer. Fig. 2 shows the example images of prostate cancers. For each prostate cancer, we cropped a square box with a patch size $P \times P$ to represent corresponding information, which was then upsampled and saved in a 224×224 resolution in 24-bit RGB JPEG format to fit the size of the 3-channel input layers of the AlexNet.

The proposed MPTL model was conducted on a desktop using a Linux OS with CPU Intel Core i7-7700@3.6 GHz, GPU NVIDIA GeForce Titan X, and 64 GB of RAM. During the standard back-propagation procedure, the momentum was set as 0.9 and weight decay was 0.005. The learning rate was set to 0.005. The weight are updated within mini-batches, and each batch includes 25 images. The training ends when the performance of the network does not improve significantly with increase of iterations. We randomly chose 80% of the dataset as the training

(50%) and validation (30%) set for transfer learning and used the remaining 20% of the images as the test set. We repeated the experiments 100 times to try different combinations of learning and test data to get the generalized results.

To enhance the robustness of the proposed MPTL model and eliminate the possibility of over-fitting, we utilized the strategy of data augmentation to increase the size of training set with various image transformations. For each prostate cancer case, we cropped every cancer region from four directions (top, bottom, left and right) in steps of 2 pixels with zero padding, and we flipped every region of interest (ROI) patch horizontally and vertically. We also considered the neighbor 2 slices of the cancer frames since cancers usually exist more than one slice. Therefore, for each patient, we cropped 21 patches to do the experiments. Fig. 3 shows the corresponding patches for a given T2w transaxial slice. Normalization was implemented to each image modality separately of each patient as in reference [31] to reduce variations among different images.

B. Performance of MPTL Model

In our model, we cropped a given size of patch around the cancer to represent the corresponding information. We evaluated recognition performance with different patch sizes, including 16×16, 24×24, 28×28 and 32×32, and recorded corresponding average and standard deviation of accuracy, recall and precision in Fig. 4. The proposed method could classify prostate cancers with relatively high accuracy (range of [0.79-0.86]) throughout the changes in patch size. In general, the best performance was obtained using a patch size of 28×28 pixels, The corresponding accuracy is 86.67%, showing the effectiveness of the proposed method for prostate cancer classification with mp-MRI images. The relative lower standard deviation shows the robustness of our method. If the patch is too small, it is difficult to describe the complete information about the prostate cancer. On the other hand, large patch size involves redundant information in the training set, which would not only degrade transfer learning performance, but also increase the calculation burdens. We fixed patch size as 28×28 in the following experiments.

We then analyzed the deep features generated from the fine-tuned MPTL model. We extracted features learned in each layer and applied the proposed image similarity loss constraint to conduct the experiments. The corresponding classification performance is shown in Fig. 5. It can be found that the accuracy increased in the first five convolutional layers. This result demonstrates the learned features represent the information of prostate cancer, therefore, may always be beneficial

to the classification. Moreover, the features calculated by the last layer are more discriminative than the features obtained from the fc6 and fc7 layer, since features that are closer to the final output layer have a higher capability of abstraction.

C. Component Analysis for MPTL Model

We compared our proposed model with eight baseline methods. The first baseline experiment is the transfer learning model without fine-tuning, which directly utilized image features learned from ImageNet to conduct experiments. The next three baseline experiments only used single MRI sequence with fine-tuning, such as T2w transaxial images, T2w sagittal images and ADC images, to carry out the prostate cancer classification. The baseline experiments 5, 6, 7 utilized two MRI parameters as input (T2w Transaxial + T2w Sagittal, T2w Transaxial + ADC and T2w Sagittal + ADC). The baseline method 8 is MPTL model with traditional softmax loss function.

Table I records the classification results of our method and eight baseline methods. Statistical significance tests are presented in Table II to allow evaluation of the comparisons of different methods. It is obvious that our MPTL model presents significantly better performance compared with these baseline experiments ($P < 0.01$). Our method achieves 4.97% higher accuracy when compared with the first baseline method. This result verified the effectiveness of fine-tuning. Actually, the fine-tuning utilizes the parameters learned from a pretrained network on a specific data set and then adjusts the parameters for the prostate cancer classification task. Therefore, this strategy can improve the accuracy significantly.

The proposed MPTL model shows better performance compared with the baseline methods 2, 3 and 4 with an improvement of 13.13%, 15.50% and 12.21% in accuracy with respect to discrimination accuracy ($P < 0.001$). One possible reason is that these baseline methods only applied a single type of MRI sequence to learn the deep features, which could not characterize the prostate cancer region well. On the contrary, our proposed method fused different MRI sequences to learn discriminative features considering that different MRI sequences represent various information of prostate cancer regions. We can see that the performance of ours was better than these of only two MRI parameters (the baseline methods 5, 6 and 7), demonstrating the necessity to use these three MRI parameters in our experiment. This result verified that different MRI sequences complement with each other to provide a good characterization of prostate cancers. Moreover, compared with the baseline 8, our method also showed higher performance. This result validates the category

similarity loss has an important role to learn discriminative features.

For more quantitative comparison, the ROC curves of eight baseline methods and ours on prostate cancer classification are shown in Fig. 6. We also conducted the DeLong test [39, 40] to compare the differences between different AUCs, which provides a confidence interval and standard error of the difference between AUCs. Our method achieves highest AUC of 0.896 as compared with these baseline methods ($P < 0.05$, Delong's test), demonstrating the statistical significance of the proposed MPTL model. We further evaluated the classification performance of two modalities in comparison with the corresponding single modality. Specifically, we conducted Delong's test for the baseline 5 (T2W Transaxial and Sagittal) vs baseline 2 (T2W Transaxial), baseline 3 (T2w Sagittal). The corresponding P value is less than 0.05, demonstrating the integration of these two MRI sequences are statistically significantly better compared than that obtained using single modality. More experiments are conducted to evaluate the baseline 6 vs Baseline 2, 4 ($P < 2.2 \times 10^{-5}$, Delong's test) and the baseline 7 vs Baseline 3, 4 ($P < 0.001$, Delong's test). These experimental results validated the effectiveness of the contributing components in our MPTL model: fine tuning procedure, integration of multiple MRI sequences and introduced image similarity constraint.

D. Comparison with Existing Methods

We evaluated the classification performance of the proposed MPTL model by comparing with the state-of-the-art prostate cancer classification methods: the radiomics based methods [8, 9, 11] and deep learning based methods [28–30]. These comparison methods have already been introduced in Section II(B).

We conducted comparison methods on our prostate cancer datasets and followed the experiment setting as in these papers to provide fair comparison. The average accuracy, recall and precision achieved by these methods and ours are shown in Table III, respectively. The deep learning based methods possess superior classification performance than the methods with radiomics features, suggesting that the deep learning based methods can learn more discriminative high-level features compared with the radiomics features. The proposed MPTL model shows better performance with an improvement of 11.05%, 11.97 and 14.97% in accuracy, 12.00%, 13.29% and 16.29% in precision, 8.60%, 9.44% and 11.25% in recall compared with methods [8, 9, 11] with significantly higher discrimination accuracy ($P < 0.005$).

Compared with the method [28], our method also shows higher recognition performance. This is because training a deep network is often complicated by convergence and overfitting issues with limited prostate cancer data. Although the methods [29, 30] tried the transfer learning to deal with the problem of prostate cancer classification, they utilized the traditional softmax loss function and treated each training image individually. In contrast, our MPTL model considered image category information in the transfer learning model. Therefore, our method calculates more accurate features for prostate cancer regions and achieves better performance in comparison with state-of-the-art prostate cancer classification methods.

V. DISCUSSION

Although our proposed MPTL model demonstrates good performance in the classification of prostate cancers, it is far from perfect. This research work focused on prostate cancer classification. Since the prediction of disease-free survival [41–44] is the ultima goal in clinical trial, we will further collect data and develop suitable models for survival prediction to promote the deep learning based radiomics. Moreover, in our proposed MPTL model, we utilized the cropped patches to represent the information of prostate cancer. To further promote our performance of prostate cancer classification, we tend to propose automatic segmentation methods to obtain the accurate prostate cancer region and then utilize these information to characterize the prostate cancer.

VI. CONCLUSION

This paper proposed a novel MPTL model to automatically classify prostate cancer. Our model transfers the knowledge of ImageNet to facilitate the learning of features of mp-MRI sequences. These transferred features are combined together to present the prostate cancer region and a fine-tuned strategy with the image category loss constraint is further introduced to enhance prostate cancer classification. This constraint enables the learned features within the same category to be close in the feature learning procedure. Therefore, the discriminative ability of the learned features is greatly improved. The superior performance of our model over the state-of-the-arts is demonstrated by a series of comparative studies. Our experimental results have shown that the proposed method is capable of classifying the stages of prostate cancer with high accuracy. Furthermore, our results have demonstrated the potential benefits of transfer learning from natural

images to medical field and may find widespread applications in cases in which annotated training datasets are limited for various practical reasons.

VII. CONFLICT OF INTEREST DISCLOSURE

The authors have no relevant conflicts of interest to disclose.

[1] Society American Cancer. Key statistics for prostate cancer `https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html/`. Accessed Jan 11, 2018.

[2] Epstein Jonathan I, Feng Zhaoyong, Trock Bruce J, Pierorazio Phillip M. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades *Eur. Urol.*. 2012;61:1019–1024.

[3] Peng Yahui, Jiang Yulei, Yang Cheng, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason scorea computer-aided diagnosis development study *Radiology.* 2013;267:787–796.

[4] Citak-Er Fusun, Vural Metin, Acar Omer, Esen Tarik, Onay Aslihan, Ozturk-Isik Esin. Final Gleason score prediction using discriminant analysis and support vector machine based on preoperative multiparametric MR imaging of prostate cancer at 3T *Biomed. Res. Int.*. 2014;2014.

[5] Niaf Emilie, Flamary Rémi, Rouviere Olivier, Lartizien Carole, Canu Stéphane. Kernel-based learning from both qualitative and quantitative labels: application to prostate cancer diagnosis based on multiparametric MR imaging *IEEE Trans. Image Process.*. 2014;23:979–991.

[6] Litjens Geert, Debats Oscar, Barentsz Jelle, Karssemeijer Nico, Huisman Henkjan. Computer-aided detection of prostate cancer in MRI *IEEE Trans. Med. Imag.*. 2014;33:1083–1092.

[7] Gillies Robert J, Kinahan Paul E, Hricak Hedvig. Radiomics: images are more than pictures, they are data *Radiology.* 2015;278:563–577.

[8] Fehr Duc, Veeraraghavan Harini, Wibmer Andreas, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images *Proc. Natl. Acad. Sci. U.S.A.* 2015;112:E6265–E6273.

[9] Vignati A, Mazzetti S, Giannini V, et al. Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness *Physics in Medicine & Biology.*

2015;60:2685.

[10] Liu Lizhi, Tian Zhiqiang, Zhang Zhenfeng, Fei Baowei. Computer-aided detection of prostate cancer with MRI: technology and applications *Acad. Radiol.*. 2016;23:1024–1046.

[11] Sobecki Piotr, Życka-Malesa Dominika, Mykhalevych Ihor, Sklinda Katarzyna, Przelaskowski Artur. MRI imaging texture features in prostate lesions classification in *EMBEC & NBC 2017*:827–830Springer 2017.

[12] Giannini Valentina, Vignati Anna, Mirasole Susanna, et al. MR-T2-weighted signal intensity: a new imaging biomarker of prostate cancer aggressiveness *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization.* 2016;4:130–134.

[13] Wang Xinggang, Yang Wei, Weinreb Jeffrey, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning *Sci. Rep.*. 2017;7:15415.

[14] Oto Aytekin, Yang Cheng, Kayhan Arda, et al. Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis *American Journal of Roentgenology.* 2011;197:1382–1390.

[15] Turkbey Baris, Shah Vijay P, Pang Yuxi, et al. Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images? *Radiology.* 2011;258:488–495.

[16] Wibmer Andreas, Hricak Hedvig, Gondo Tatsuo, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores *European radiology.* 2015;25:2840–2850.

[17] Källén Hanna, Molin Jesper, Heyden Anders, Lundström Claes, Åström Kalle. Towards grading gleason score using generically trained deep convolutional neural networks in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*:1163–1167IEEE 2016.

[18] LeCun Yann, Bengio Yoshua, Hinton Geoffrey. Deep learning *Nature.* 2015;521:436–444.

[19] Esteva Andre, Kuprel Brett, Novoa Roberto A, et al. Dermatologist-level classification of skin cancer with deep neural networks *Nature.* 2017;542:115–118.

[20] Albarqouni Shadi, Baur Christoph, Achilles Felix, Belagiannis Vasileios, Demirci Stefanie, Navab Nassir. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images *IEEE Trans. Med. Imag.*. 2016;35:1313–1321.

[21] Yuan Yixuan, Meng Max Q-H. Deep learning for polyp recognition in wireless capsule endoscopy images *Med. Phys.*. 2017;44:1379–1389.

[22] Pan Sinno Jialin, Yang Qiang. A survey on transfer learning *IEEE Trans. Knowl. Data Eng.*. 2010;22:1345–1359.

[23] Raina Rajat, Battle Alexis, Lee Honglak, Packer Benjamin, Ng Andrew Y. Self-taught learning: transfer learning from unlabeled data in *Proceedings of the 24th international conference on Machine learning*:759–766ACM 2007.

[24] Vargas Hebert Alberto, Wassberg Cecilia, Akin Oguz, Hricak Hedvig. MR imaging of treated prostate cancer *Radiology.* 2012;262:26–42.

[25] Li Cheng, Liu Haifeng, Cai Deng. Active learning on manifolds *Neurocomputing.* 2014;123:398–405.

[26] Reda Islam, Shalaby Ahmed, Khalifa Fahmi, et al. Computer-aided diagnostic tool for early detection of prostate cancer in *Image Processing (ICIP), 2016 IEEE International Conference on*:2668–2672IEEE 2016.

[27] Tsehay Yohannes, Lay Nathan, Wang Xiaosong, et al. Biopsy-guided learning with deep convolutional neural networks for Prostate Cancer detection on multiparametric MRI in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*:642–645IEEE 2017.

[28] Kiraly Atilla P, Nader Clement Abi, Tuysuzoglu Ahmet, et al. Deep Convolutional Encoder-Decoders for Prostate Cancer Detection and Classification in *International Conference on Medical Image Computing and Computer-Assisted Intervention*:489–497Springer 2017.

[29] Chen Quan, Xu Xiang, Hu Shiliang, Li Xiao, Zou Qing, Li Yunpeng. A Transfer learning approach for classification of clinical significant prostate cancers from mpMRI scans in *SPIE Medical Imaging*:101344F–101344FInternational Society for Optics and Photonics 2017.

[30] Le Minh Hung, Chen Jingyu, Wang Liang, et al. Automated diagnosis of prostate cancer in multiparametric MRI based on multimodal convolutional neural networks *Phys. Med. Biol.*. 2017;62:6497.

[31] Song Yang, Zhang Yu-Dong, Yan Xu, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI *Journal of Magnetic Resonance Imaging.* 2018.

[32] Zhang Ruikai, Zheng Yali, Mak Tony Wing Chung, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain *IEEE J. Biomed. Health Inform.*. 2017;21:41–47.

[33] Cheng Phillip M, Malhi Harshawn S. Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images *J. Digit. Imaging.* 2017;30:234–243.

[34] Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E. Imagenet classification with deep convolutional neural networks in *Proc. Int. Conf. Neural Infor. Proces.*:1097–1105 2012.

[35] Szegedy Christian, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*:1–9 2015.

[36] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556.* 2014.

[37] Yang Xin, Wang Zhiwei, Liu Chaoyue, et al. Joint Detection and Diagnosis of Prostate Cancer in Multi-parametric MRI Based on Multimodal Convolutional Neural Networks in *International Conference on Medical Image Computing and Computer-Assisted Intervention*:426–434Springer 2017.

[38] Bottou Léon. Stochastic gradient descent tricks in *Neural networks: Tricks of the trade*:421–436Springer 2012.

[39] DeLong Elizabeth R, DeLong David M, Clarke-Pearson Daniel L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach *Biometrics.* 1988:837–845.

[40] Yin Jieyun, Liu Hongliang, Liu Zhensheng, et al. Genetic variants in fanconi anemia pathway genes BRCA2 and FANCA predict melanoma survival *Journal of Investigative Dermatology.* 2015;135:542–550.

[41] Gnep Khémara, Fargeas Auréline, Gutiérrez-Carvajal Ricardo E, et al. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer *Journal of Magnetic Resonance Imaging.* 2017;45:103–117.

[42] Guinney Justin, Wang Tao, Laajala Teemu D, et al. Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data *The Lancet Oncology.* 2017;18:132–142.

[43] Kang J, Doucette CW, Zhang H. Machine Learning to Predict Postradical Prostatectomy Pathology Outcomes in Intermediate Risk Prostate Cancer *International Journal of Radiation Oncology Biology Physics.* 2017;99:E245–E246.

[44] Nezhad Milad Zafar, Sadati Najibesadat, Yang Kai, Zhu Dongxiao. A Deep Active Survival Analysis approach for precision treatment recommendations: Application of prostate cancer *Expert Systems with Applications.* 2019;115:16–26.

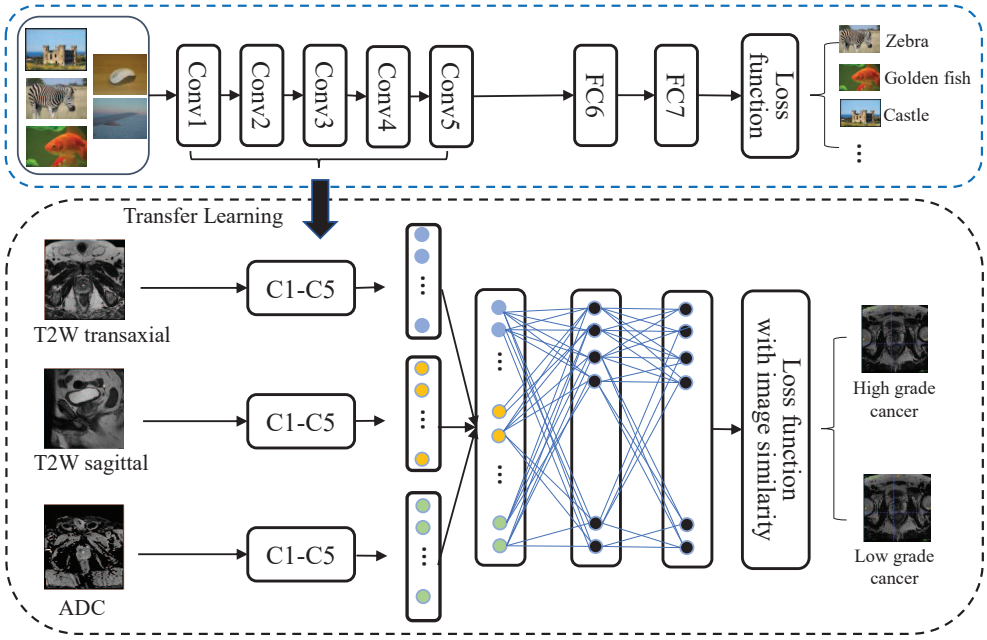

FIG. 1. Workflow of our proposed MPTL model for diagnosis of prostate cancer. C represents convolutional layer while FC represents the fully connected layer.

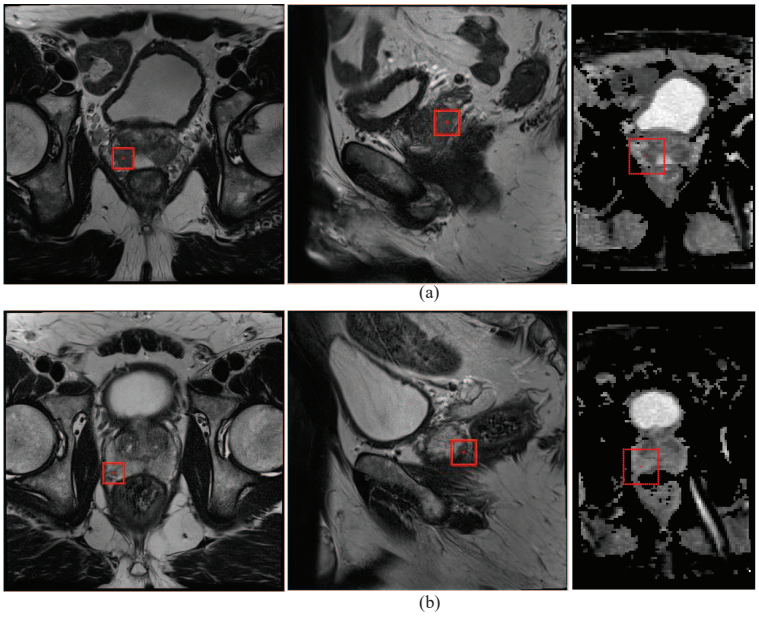

FIG. 2. Illustrations of images with prostate cancers and the corresponding cropped 28×28 patches. (a) Prostate cancer with low grade, (b) Prostate cancer with high grade. The different columns represent the T2w transaxial image, T2w sagittal image and ADC map, respectively. The red squares with a size of 28×28 represent cancer region while the red points in the patch center represent the annotated cancer region.

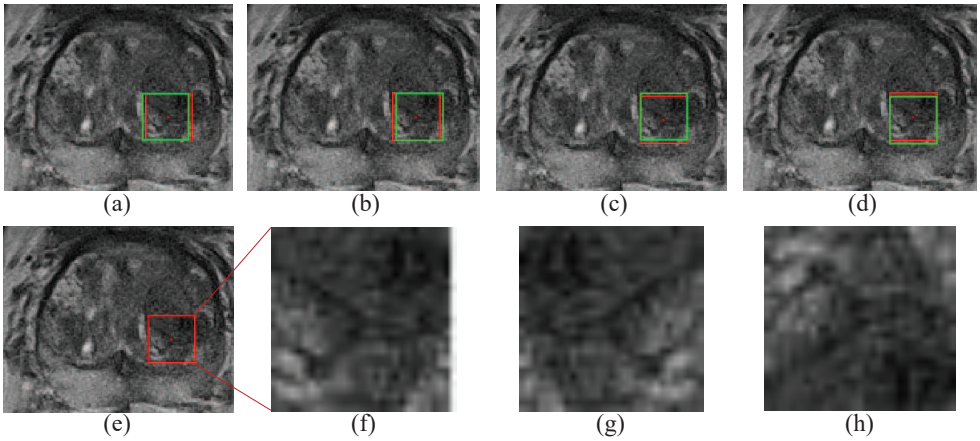

FIG. 3. Illustrations of cropped patches with prostate cancers. (a-d) The red square represents the cropped 28×28 patch while the green squares show the cropped cancer region from four directions (top, bottom, left and right) in steps of 2 pixels with zero padding. (e-f) show the cropped cancer patch and (g-h) show the horizontal and vertical flipped cancer patches.

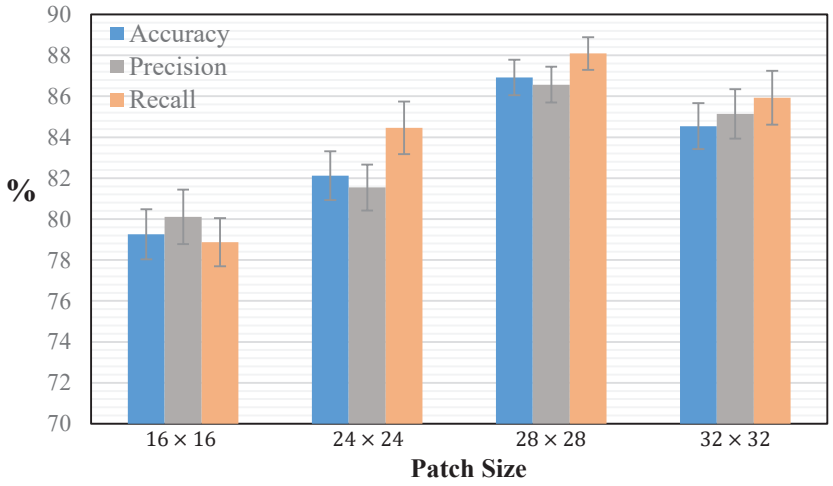

FIG. 4. Diagnosis performance with different patch sizes. X-axis represents different patch sizes while Y-axis shows the corresponding classification performance.

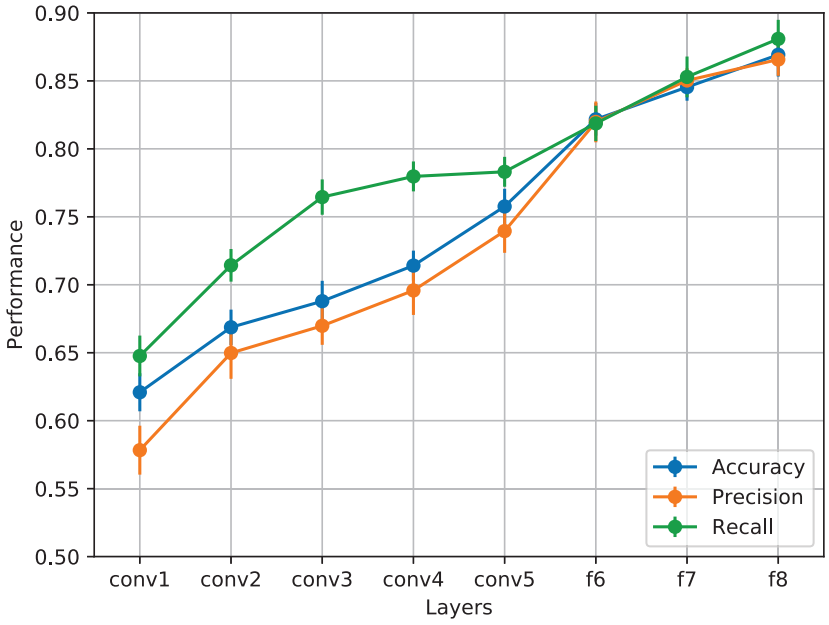

FIG. 5. Diagnosis performance with features from different layers. X-axis represents different layers in the proposed MPTL model while Y-axis shows the corresponding classification performance.

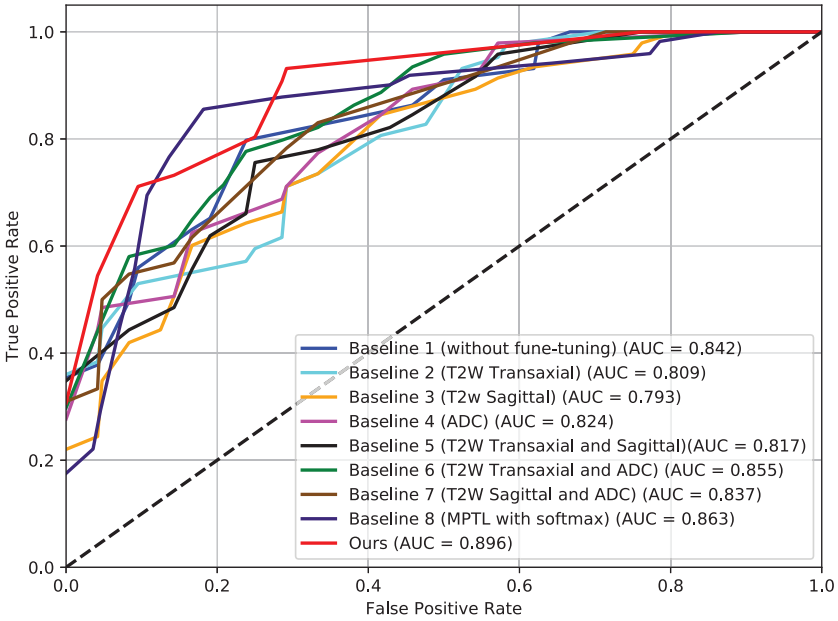

FIG. 6. ROC curves for different baseline methods.

TABLE I. Comparison results of baseline methods.

	Acc (%)	**Pre** (%)	**Rec** (%)	**F1** (%)
Baseline 1 (without fune-tuning)	81.95 ± 1.83	81.85 ± 1.71	82.57 ± 1.62	82.26 ± 1.84
Baseline 2 (T2W Transaxial)	73.79 ± 1.36	71.42 ± 1.42	79.39 ± 1.24	75.44 ± 1.79
Baseline 3 (T2w Sagittal)	71.42 ± 1.14	72.14 ± 1.24	73.79 ± 1.47	72.67 ± 1.48
Baseline 4 (ADC)	74.71 ± 1.28	73.85 ± 1.71	78.31 ± 1.13	76.06 ± 1.35
Baseline 5 (T2W Transaxial and Sagittal)	74.95 ± 1.06	74.14 ± 1.28	78.56 ± 1.26	75.86 ± 1.12
Baseline 6 (T2W Transaxial and ADC)	77.78 ± 0.93	76.18 ± 1.14	80.97 ± 1.12	78.19 ± 1.06
Baseline 7 (T2W Sagittal and ADC)	75.59 ± 1.31	75.52 ± 1.27	77.85 ± 1.39	76.15 ± 1.29
Baseline 8 (MPTL with softmax)	84.57 ± 1.07	84.12 ± 1.48	85.98 ± 1.22	85.09 ± 1.13
Ours	86.92 ± 1.09	86.57 ± 0.85	88.09 ± 1.18	87.86 ± 1.23

TABLE II. Statistical significant (p-values) of accuracy for comparing baseline methods and our method.

	Baseline 1 *vs* Ours	Baseline 2 *vs* Ours	Baseline 3 *vs* Ours	Baseline 4 *vs* Ours
p-values	0.003	<0.001	<0.001	<0.001
	Baseline 5 *vs* Ours	Baseline 6 *vs* Ours	Baseline 7 *vs* Ours	Baseline 8 *vs* Ours
p-values	<0.001	<0.001	<0.001	0.005

TABLE III. Comparison results of existing methods for prostate cancer classification.

	Acc (%)	**Pre** (%)	**Rec** (%)	**F1** (%)
Fehr et al. [8]	75.87 ± 1.58	74.57 ± 0.79	79.49 ± 1.49	76.14 ± 0.95
Vignati et al. [9]	74.95 ± 1.17	73.28 ± 1.24	78.65 ± 1.31	75.48 ± 0.71
Sobecki et al. [11]	71.95 ± 1.63	70.28 ± 1.14	76.84 ± 1.56	73.66 ± 1.13
Kiraly et al. [28]	77.78 ± 1.26	76.58 ± 0.85	82.19 ± 1.57	79.88 ± 1.28
Chen et al. [29]	80.09 ± 1.58	78.95 ± 1.57	83.96 ± 1.34	81.61 ± 1.63
Le et al. [30]	82.09 ± 1.76	82.27 ± 1.14	82.88 ± 1.52	82.34 ± 1.38
Ours	86.92 ± 1.09	86.57 ± 0.85	88.09 ± 1.18	87.86 ± 1.23