Prostate Cancer Classification on VERDICT DW-MRI Using Convolutional Neural Networks

Eleni Chiou[1,2], Francesco Giganti[3,4], Elisenda Bonet-Carne[5], Shonit Punwani[5], Iasonas Kokkinos[1], and Eleftheria Panagiotaki[1,2]

[1] Department of Computer Science, UCL, London, UK
[2] Centre for Medical Image Computing, UCL, London, UK
[3] Department of Radiology, UCLH NHS Foundation Trust, London, UK
[4] Division of Surgery & Interventional Science, UCL, London, UK
[5] Centre for Medical Imaging, Division of Medicine, UCL, London, UK

Abstract. Currently, non-invasive imaging techniques such as magnetic resonance imaging (MRI) are emerging as powerful diagnostic tools for prostate cancer (PCa) characterization. This paper focuses on automated PCa classification on VERDICT (Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumors) diffusion weighted (DW)-MRI, which is a non-invasive microstructural imaging technique that comprises a rich imaging protocol and a tissue computational model to map in vivo histological indices. The contribution of the paper is two fold. Firstly, we investigate the potential of automated, model-free PCa classification on raw VERDICT DW-MRI. Secondly, we attempt to adapt and evaluate novel fully convolutional neural networks (FCNNs) for PCa characterization. We present two neural network architectures that adapt U-Net and ResNet-18 to the PCa classification problem. We train the networks end-to-end on DW-MRI data and evaluate the diagnostic performance employing a 10-fold cross validation approach using data acquired from 103 patients. ResNet-18 outperforms U-Net with an average AUC of 86.7%. Our results show promise for the utilization of raw VERDICT DW-MRI data and FCNNs for automating the PCa diagnostic pathway.

Keywords: VERDICT MRI · Prostate Cancer Classification · Convolutional Neural Networks.

1 Introduction

Prostate cancer (PCa) is the second most common cancer among men worldwide [1]. Early diagnosis and treatment are crucial to decrease the mortality rate in patients. Thus, the development of reliable and automated diagnostic tools is imperative. Currently, multi-parametric (mp)-magnetic resonance imaging (MRI), which consists of T2-weighted (T2W)-MRI, diffusion-weighted (DW)-MRI and dynamic contrast enhanced (DCE)-MRI, has become a useful tool for non-invasive PCa diagnosis. Moreover, VERDICT (Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumors) DW-MRI, which is an advanced microstructural imaging technique for cancer characterization has currently been proposed as an additional, powerful diagnostic tool. However, radiological interpretation of different MRI sequences for PCa characterization remains a complex and time-consuming task.

Several methods for automated diagnosis of PCa on mp-MRI have been proposed to improve diagnostic accuracy and speed up the decision-making process. State-of-the-art methods use deep learning to classify and localize PCa on mp-MRI. For example, Kiraly et al. proposed a multi-channel image-to-image convolutional encoder-decoder to classify lesions achieving an area under the curve (AUC) of 83% [2]. Mehrtash et al. used a 3D convolutional neural network for automated detection of PCa on mp-MRI and reported an AUC of 80% [3]. However, these studies share a common limitation, i.e., they rely on mp-MRI data, which despite its merits has been shown to have low specificity [4].

DW-MRI has been demonstrated to be the most important component of mp-MRI compared to T2W-MRI and DCE-MRI due to its high sensitivity to microstructural changes related to cancer [5], [6]. However, mp-MRI studies use DW-MRI in its simplest form by deriving the ADC map. This simplified model of water diffusion lacks biological specificity as it fails to discriminate the variety of histological changes that occur in cancer [7]. VERDICT DW-MRI improves on ADC maps by modelling directly the underlying microstructure [8], [9]. More specifically, VERDICT combines an optimized DW-MRI acquisition protocol with a mathematical model to estimate and map microstructural features such as cell size, density, and vascular volume fraction, all of which change in cancer.

In this paper, we first aim to investigate the potential of model-free PCa characterization using the raw DW-MRI data from the VERDICT acquisition. Second, we attempt to adapt and evaluate fully convolutional neural networks (FCNNs) for automated characterization of PCa on VERDICT DW-MRI data.

2   Methods

2.1   VERDICT DW-MRI data

In this study we use VERDICT DW-MRI data from 103 patients. VERDICT DW-MRI images (Fig. 1) were acquired with pulsed-gradient spin-echo sequence (PGSE) using an optimised imaging protocol for VERDICT prostate characterization with 5 b-values $(90, 500, 1500, 2000, 3000 \, \text{s/mm}^2)$ in 3 orthogonal directions, on a 3T scanner (Achieva, Philips Healthcare, NL) [10]. Also, images with $b = 0 \, \text{s/mm}^2$ were acquired before each b-value acquisition. The DW-MRI sequence was acquired with a voxel size of $1.25 \times 1.25 \times 5 \, \text{mm}^3$, 5 mm slice thickness, 14 slices, a field of view of $220 \times 220 \, \text{mm}^2$ and the images were reconstructed to a 176×176 matrix size. The data was registered using rigid registration [11]. A dedicated radiologist highly experienced in prostate mp-MRI reporting (reporting more than 1000 scans per year) contoured malignant and benign lesions on the registered VERDICT DW-MRI using mp-MRI for guidance.

2.2   Fully Convolutional Neural Networks

We formulate the problem as pixel-wise classification and use FCNNs trained end-to-end, on DW-MRI data. FCNNs have shown great success on pixel-wise

(a) $b = 90\,\mathrm{s/mm^2}$ (b) $b = 500\,\mathrm{s/mm^2}$ (c) $b = 1500\,\mathrm{s/mm^2}$

(d) $b = 2000\,\mathrm{s/mm^2}$ (e) $b = 3000\,\mathrm{s/mm^2}$
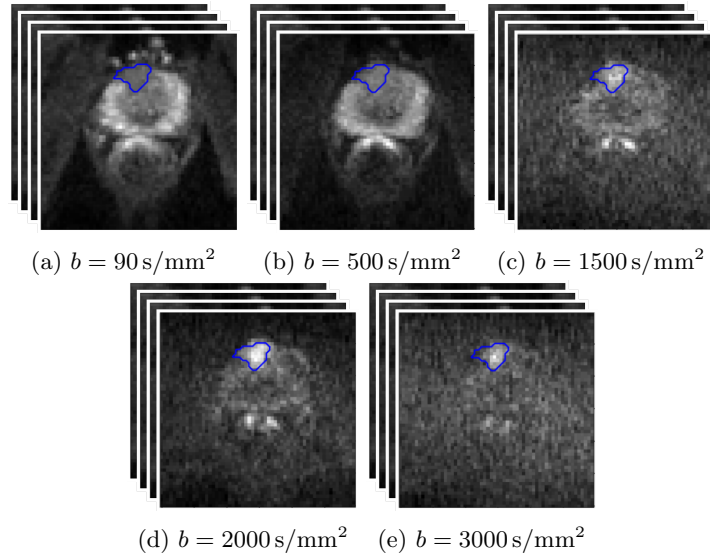
Fig. 1: VERDICT DW-MRI data acquired with 5 b-values in 3 orthogonal directions. Cancerous regions (noted in blue) are seen as a focus of high signal intensity on DW-MRI of $b = 2000, 3000\,\mathrm{s/mm^2}$ and as a focus of low signal intensity on the corresponding $b = 90\,\mathrm{s/mm^2}$ image.

classification tasks [12], [13], [14]. They adapt standard CNNs to the pixel-wise classification problem by converting fully connected layers to convolutional layers and adding upsampling layers to reach the original image resolution. They get as input the entire image and produce pixel-wise class probability maps. We implement and evaluate two encoder-decoder FCNNs. Encoder-decoder networks extract low resolution feature maps in the encoder module and gradually map them to full input resolution feature maps in the corresponding decoder module [13], [14], [15].

Objective function. We consider two classes (malignant, benign/normal) and perform pixel-wise classification on predefined regions of interest (ROIs) on DW-MRI data. We define a label set $\mathcal{L} = \{0, 1, 2\}$, where 0 corresponds to benign/normal, 1 to malignant and 2 to ambiguous. Let $\mathbf{I} \in \mathbb{R}^{176 \times 176 \times 20}$ be a 20-channel DW-MRI image and $\mathbf{L} \in \mathcal{L}^{176 \times 176}$ the corresponding labelling. Also, let $\mathbf{P} \in [0, 1]^{176 \times 176}$ be the pixel-wise probability map indicating the probability of each pixel to belong to class 1. We train the networks using pixel-wise cross-entropy on pixels $j = \{j : l_j \neq 2$, where l_j the label of pixel $j\}$. The cross-entropy loss is defined as

$$\mathrm{CE} = -\sum_j l_j \log(p_j) + (1 - l_j) \log(1 - p_j), \tag{1}$$

where p_j is the probability of pixel j belonging to class 1 and l_j is the class label of pixel j.

Table 1: MRI-UNet. The encoder module includes 3 encoder blocks (encBlock). Each encBlock consists of a convolutional layer (conv) followed by batch normalization (BN), a rectified-linear unit (ReLU) and a max pooling operation (pool). The decoder module includes 3 decoder blocks (decBlock). Each decBlock consists of a convolutional layer followed by BN, a ReLU and a transposed convolution (convTransp). The last block (outBlock) consists of 2 convolutional layers. The last convolutional layer is followed by a pixel-wise softmax.

	Layer	Description
inBlock	input	64x64 DW-MRI images
encBlock1	conv1	3x3, 64, stride 1, BN, ReLU
	pool1	2x2 maxpooling, stride 2
encBlock2	conv2	3x3, 128, stride 1, BN, ReLU
	pool2	2x2 maxpooling, stride 2
encBlock3	conv3	3x3, 256, stride 1, BN, ReLU
	pool3	2x2 maxpooling, stride 2
decBlock1	conv4	3x3, 256, stride 1, BN, ReLU
	convTransp1	2x2, 256, stride 2
dencBlock2	conv5	3x3, 256, stride 1, BN, ReLU
	convTransp2	2x2, 128, stride 2
dencBlock3	conv6	3x3, 128, stride 1, BN, ReLU
	convTransp3	2x2, 64, stride 2
outBlock	conv7	3x3, 64, stride 1, BN, ReLU
	conv8	1x1, 2, stride 1, softmax

Table 2: MRI-ResNet. The encoder module includes 3 encoder blocks (encBlock). Each encBlock consists of 2 convolutional layers (conv) each followed by batch normalization (BN) and a rectified-linear unit (ReLU). The decoder module includes 3 decoder blocks (decBlock). Each decBlock consists of a bilinear upsapmling operation (bilUp) followed by 2 convolutional layers. The last convolutional layer is followed by a pixel-wise softmax.

	Layer	Description
inBlock	input	64x64 DW-MRI images
	conv1	7x7, 64, stride 2, BN, ReLU
encBlock1	conv2	3x3, 64, stride 1, BN, ReLU
	conv3	3x3, 64, stride 1, BN, ReLU
encBlock2	conv4	3x3, 128, stride 2, BN, ReLU
	conv5	3x3, 128, stride 1, BN, ReLU
encBlock3	conv6	3x3, 128, stride 2, BN, ReLU
	conv7	3x3, 128, stride 1, BN, ReLU
decBlock1	bilUp1	scale factor 2
	conv8	1x1, 128, stride 1, BN, ReLU
	conv9	3x3, 64, stride 1, BN, ReLU
decBlock2	bilUp2	scale factor 2
	conv10	1x1, 64, stride 1, BN, ReLU
	conv11	3x3, 64, stride 1, BN, ReLU
decBlock3	bilUp3	scale factor 2
	conv12	1x1, 64, stride 1, BN, ReLU
	conv13	3x3, 2, stride 1, BN, softmax

Network Architectures. The first network (MRI-UNet) (Table 1) is based on the U-Net architecture proposed in [13]. U-Net consists of an encoder module and a symmetric decoder module. MRI-UNet has fewer convolutional layers to avoid overfitting. The encoder module includes 3 encoder blocks (encBlock). Each encBlock consists of a convolutional layer followed by batch normalization (BN) [16], a rectified-linear unit (ReLU) [17] and a 2x2 max pooling operation with stride 2. Each convolutional layer performs 2D convolutions of the input maps with 3x3 kernels. The decoder module includes 3 decoder blocks (decBlock). Each decBlock consists of a convolutional layer followed by BN, a ReLU and a 2x2 transposed convolution with stride 2 to upsample low resolution feature maps. Concatenation of the upsampled feature maps with the corresponding encoder feature maps is performed before the convolutional layers. Each convolutional layer performs 2D convolutions of the input with 3x3 kernels. The last convolutional layer is followed by a pixel-wise softmax which provides class probability maps.

The second network (MRI-ResNet) has also an encoder-decoder structure (Table 2). The encoder module is similar to the ResNet-18 network proposed in [18]. We remove the max pooling layer in the beginning of the network and the global average pooling layer at the end of the network. Also, we replace the last fully-connected layer with a convolutional layer and decrease the number of convolutional layers. The encoder module includes 3 encBlock. Each encBlock consists of 2 convolutional layers each followed by BN and a ReLU. The decoder module is similar to the one proposed in [15] and has 3 decBlock. Each decBlock consists of a bilinear upsapmling operation followed by 2 convolutional layers. The low resolution feature maps are bilinear upsampled by a factor of 2 and then concatenated with the corresponding encoder feature maps. Then, a 1x1 convolutional layer followed by a 3x3 convolutional layer are applied to reduce the number of feature maps and refine the features. The last convolutional layer is followed by a pixel-wise softmax.

Training settings. We implement both networks using Pytorch [19]. We employ a 10-fold cross validation (CV) approach to train and test the networks. We repeat each 10-fold CV 5 times. We train the networks for 200 epochs and select the model which has the smallest loss on a validation set (20% of the training set). We use stochastic gradient descent (SGD) with a mini-batch size of 32, a constant learning rate of 1e-5, a momentum of 0.9 and a weight decay of 1e-3.

Evaluation metrics. We evaluate the binary pixel-wise classification using average sensitivity, specificity, AUC and precision. Sensitivity, specificity and precision are defined as

- sensitivity $= \frac{TP}{P}$, where TP is the number of true positive pixels and P is the number of positive pixels.
- specificity $= \frac{TN}{N}$, where TN is the number of true negative pixels and N is the number of negative pixels.
- precision $= \frac{TP}{FP+TP}$, where FP is the number of false positive pixels.

Table 3: Average AUC, sensitivity, specificity, precision of MRI-UNet and MRI-ResNet when evaluation is performed on i) malignant and benign ROIs (malignant vs benign ROIs) and ii) the entire image (malignant vs all).

Networks	Regions	AUC	sensitivity	specificity	precision
MRI-UNet	malignant vs benign ROIs	85.7%	75.7%	75.4%	86.2%
	malignant vs all	74.1%	75.6%	47.6%	2.0%
MRI-ResNet	malignant vs benign ROIs	86.7%	71.8%	83.3%	90.0%
	malignant vs all	71.9%	71.6%	57.1%	4.7%

Table 4: Average AUC, sensitivity, specificity, precision of MRI-UNet and MRI-ResNet when evaluation is performed on i) malignant and benign ROIs (malignant vs benign ROIs) and ii) the entire image (malignant vs all) when we use additional negative labelled ROIs.

Networks	Regions	AUC	sensitivity	specificity	precision
MRI-UNet	malignant vs benign ROIs	89.0%	82.9%	77.9%	88.8%
	malignant vs all	94.2%	82.7%	91.2%	13.4%
MRI-ResNet	malignant vs benign ROIs	87.6%	86.4%	72.6%	86.7%
	malignant vs all	94.0%	86.4%	88.8%	11.2%

3   Results

We perform two different experiments and report the results.

1st experiment. In the first experiment we train the networks on predefined malignant and benign/normal ROIs and ignore the rest of the pixels. Then, we evaluate the classification performance of the networks on i) predefined malignant and benign/normal ROIs and ii) the entire image (Table 3). In the second case regions which are not labelled as malignant are considered as benign/normal. Using MRI-ResNet results in slightly improved performance (AUC of 86.7%). Figure 2 shows the receiver operating characteristic (ROC) curves of MRI-UNet and MRI-ResNet when evaluation is performed on i) predefined malignant and benign/normal ROIs and ii) the entire image.

2nd experiment. In the previous experiment the dataset is highly unbalanced since the number of malignant ROIs is higher than the number of benign/normal ROIs. To address this issue, we increase the number of negative labelled ROIs by randomly selecting and adding normal/background ROIs to the training set. Table 4 shows the classification performance of the networks when we increase the number of negative labelled ROIs. Using additional negative labelled ROIs improves AUC when classification is performed on the predefined regions or the entire image. Figure 3 shows the ROC curves of MRI-UNet and MRI-ResNet when evaluation is performed on i) predefined malignant and benign ROIs and ii) the entire image.

The results show that MRI-ResNet achieves an AUC of 86.7% in classifying benign and malignant ROIs. A recent study which uses an encoder-decoder network for classification of malignant and benign regions on the PROSTATEx dataset [20] reports a maximum AUC of 83.4% using mp-MRI data and a maximum AUC of 80.4% using the ADC map derived from the DW-MRI data [2].

Comparison of our results to the results of this study shows that VERDICT DW-MRI data combined with FCNNs give comparable performance and could be used as an alternative automated diagnostic tool for PCa classification. However, this comparison is limited due to the difference in the two datasets.
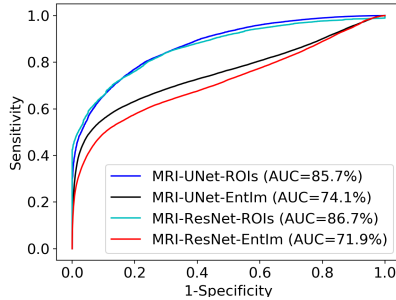

Fig. 2: 1st experiment. Receiver operating characteristic (ROC) curves of MRI-UNet and MRI-ResNet when evaluation is performed on i) predefined malignant and benign ROIs and ii) the entire image (EntIm).
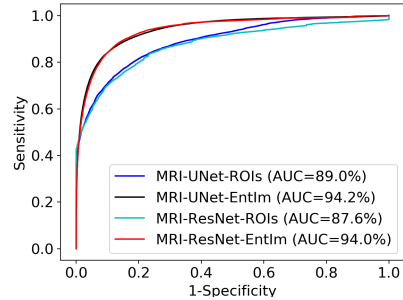
Fig. 3: 2nd experiment. Receiver operating characteristic (ROC) curves of MRI-UNet and MRI-ResNet when evaluation is performed on i) predefined malignant and benign ROIs and ii) the entire image (EntIm).

4   Conclusion

We investigate the potential of model-free PCa classification on the raw VERDICT DW-MRI data using FCNNs. For this purpose, we adapt and evaluate two FCNN architectures. Previous studies are based on mp-MRI data to provide an automated solution for PCa classification. In this study, we use richer DW-MRI compared to DW-MRI from mp-MRI acquisitions, acquired for 5 b-values in 3 orthogonal directions to train and evaluate the FCNNs. MRI-ResNet behaves better than MRI-UNet achieving an AUC of 86.7% in classifying malignant and benign regions. The results indicate that VERDICT DW-MRI data combined with FCNNs show promise as an alternative diagnostic tool for PCa classification.

Future work will address the PCa classification problem in the entire image. The proposed networks are not suitable for PCa classification in the entire image due to foreground-background class imbalance characterizing the data. To address this issue we plan to include an additional step for prostate segmentation so as to limit the analysis on the prostate region. Finally, we plan to investigate techniques to provide visual explanations of decision from FCNNs models, which is crucial for medical diagnosis applications.

Acknowledgments

References

1. Torre, L.A., et al.: Global cancer statistics, 2012. CA: Cancer J. Clin. (2015)
2. Kiraly, A.P., et al.: Deep convolutional encoder-decoders for prostate cancer detection and classification. In: MICCAI (2017)
3. Mehrtash, A., et al.: Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks. In: Proc. SPIE Int. Soc. Opt. Eng. (2017)
4. Ahmed, H.U., et al.: Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. The Lancet (2017)
5. Isebaert, S., et al.: Multiparametric MRI for prostate cancer localization in correlation to wholemount histopathology. J. Magn. Reson. Imaging (2013)
6. Metzger, G.J., et al.: Detection of prostate cancer: Quantitative multiparametric MR imaging models developed using registered correlative histopathology. Radiol. (2016)
7. Bourne, R., et al.: Limitations and prospects for diffusion-weighted MRI of the prostate. Diagnostics (2016)
8. Panagiotaki, E., et al.: Noninvasive quantification of solid tumor microstructure using VERDICT MRI. Cancer Res. (2014)
9. Panagiotaki, E., et al.: Microstructural characterization of normal and malignant human prostate tissue with vascular, extracellular, and restricted diffusion for cytometry in tumours magnetic resonance imaging. Invest. Radiol. (2015)
10. Panagiotaki, E., et al.: Optimised VERDICT MRI protocol for prostate cancer characterisation. In: ISMRM (2015)
11. Ourselin, S., et al.: Reconstructing a 3D structure from serial histological sections. Image Vis. Comput. (2001)
12. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
13. Ronneberger, O., et al.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
14. Badrinarayanan, V., et al.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. (2017)
15. Chen, L.C., et al.: Encoder-decoder with atrous separable convolution for semantic image segmentation. ArXiv (2018)
16. Ioffe, S., et al.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
17. Nair, V., et al.: Rectified linear units improve restricted boltzmann machines. In: ICML (2010)
18. He, K., et al.: Deep residual learning for image recognition. In: CVPR (2016)
19. Paszke, A., et al.: Automatic differentiation in pytorch. In: Autodiff Workshop, NIPS (2017)
20. Litjens, G., et al: Computer-aided detection of prostate cancer in MRI. IEEE Trans. Med. Imaging (2014)