Joint estimation of relaxation and diffusion tissue parameters for prostate cancer grading with relaxation-VERDICT MRI

Marco Palombo[1*], Vanya Valindria[1*], Saurabh Singh[2], Eleni Chiou[1], Francesco Giganti[3,4], Hayley Pye[5], Hayley C. Whitaker[5], David Atkinson[2], Shonit Punwani[2], Daniel C. Alexander[1], Eleftheria Panagiotaki[1].

Affiliations:

1 Centre for Medical Image Computing, Department of Computer Science, University College London, London (UK)

2 Centre for Medical Imaging, University College London, London (UK)

3 Division of Surgery and Interventional Science, University College London, London (UK)

4 Department of Radiology, University College London Hospital NHS Foundation Trust, London (UK)

5 Molecular Diagnostics and Therapeutics Group, Division of Surgery & Interventional Science, University College London, London, (UK)

*M.P. and V.V. contributed equally to this work

Sources of funding: this work was supported by EP/N021967/1, EP/R006032/1 and by Prostate Cancer UK: Targeted Call 2014: Translational Research St.2, project reference PG14-018-TR2. M.P. is supported by the UKRI Future Leaders Fellowship MR/T020296/1.

Correspondence to: Marco Palombo, PhD; Centre for Medical Image Computing, Department of Computer Science, University College London, 90 High Holborn, London WC1V 6LJ (UK). E-mail: mrc.palombo@gmail.com.

Running head: Non-invasive Gleason scoring with rVERDICT MRI

ABSTRACT

Objectives. The Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumours (VERDICT) is a non-invasive MRI technique that has shown promising results in clinical settings discriminating normal from malignant tissue and Gleason grade 3+3 from 3+4. However, VERDICT currently does not account for the inherent relaxation properties of the tissue, whose quantification could provide additional information and enhance its diagnostic power. The aim of this work is to introduce relaxation-VERDICT (rVERDICT) for prostate, a model for the joint estimation of diffusion-based parameters (e.g. the intracellular volume fraction, f_{ic}) and relaxation times (e.g. T2) from a VERDICT MRI acquisition; and to evaluate its repeatability and diagnostic utility for differentiating Gleason grades 3 and 4.

Materials and Methods. 72 men with suspected prostate cancer (PCa) or undergoing active surveillance were recruited. All men underwent multiparametric MRI (mp-MRI) and VERDICT MRI. Deep neural network was used for ultra-fast fitting of the rVERDICT parameters. 44 men underwent targeted biopsy, which enabled assessment of rVERDICT parameters in differentiating Gleason grades measured with accuracy, F1-score and Cohen's kappa of a convolutional neural network classifier. To assess the repeatability of the new model, five men were imaged twice.

Results. The rVERDICT intracellular volume fraction f_{ic} discriminated between Gleason grades (5-class classification) with accuracy, F1-score and kappa 8, 7 and 3 percentage points higher than classic VERDICT, and 12, 13 and 24 percentage points higher than the Apparent Diffusion Coefficient (ADC) from mp-MRI. Repeatability of rVERDICT parameters was high (R^2=0.74–0.99, coefficient of variation=1%–10% and intraclass correlation coefficient=78%-98%). T2 values estimated with rVERDICT were not significantly different from those estimated with an independent multi-TE acquisition ($p>0.05$). The deep neural network fitting approach provided stable fitting of all the rVERDICT parameters with dramatic reduction of the processing time (~35 seconds *vs* ~15 minutes using classic VERDICT), enabling on-the-fly rVERDICT map generation.

Conclusions. The new rVERDICT allows for robust and ultra-fast microstructural estimation of diffusion and relaxation properties of PCa and enables Gleason scoring.

Key words: microstructure imaging, diffusion MRI, relaxation MRI, multiparametric MRI, biophysical modelling, deep learning, prostate cancer microstructure, Gleason grade, prostate cancer diagnosis, prostate cancer grading.

INTRODUCTION

As for many cancers, prostate cancer (PCa) definitive diagnosis relies on biopsies. This invasive procedure can have serious side effects, such as infection, bleeding and urinary retention, significantly impacting quality of life. Recent advances in medical imaging have played a key role in improving PCa detection. For instance, multi-parametric MRI (mp-MRI), consisting of T2-weighted, diffusion-weighted and dynamic contrast-enhanced imaging sequences has been incorporated into the National Institute for Health and Care Excellence (NICE) guidelines for PCa diagnosis.(1) However, whilst mp-MRI has 90% sensitivity for detection of significant cancer, it's specificity is moderate at 50%;(2) resulting in 1 in 2 men still needing to undergo an unnecessary biopsy.(3) Significant cancer is generally defined by the presence of Gleason pattern 4 tumour within a biopsy. Reliably separating lesions on mp-MRI that contain Gleason pattern 4 disease from those with non-significant cancer (Gleason 3+3) or no cancer remains an unmet clinical need.

To address this, microstructure imaging techniques based on diffusion-weighted MRI (DW-MRI)(4-6) offer sensitivity and specificity to microstructure changes that correspond to different Gleason grades, well above the simple apparent diffusion coefficient (ADC), conventionally acquired as part of standard mp-MRI protocols.(7) For example, the Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumours (VERDICT) technique(5, 8) was one of the first non-invasive imaging techniques to show histological specificity both ex-vivo and in-vivo, as well as in preclinical and clinical studies.(5, 8-10) Preliminary results from the clinical trial INNOVATE(11) reveal that the VERDICT intracellular volume fraction (f_{ic}) can discriminate between Gleason 3+3 and 3+4 lesions.

However, VERDICT is currently limited in estimating only diffusion parameters without accounting for the inherent relaxation properties of the tissue.(12-19) This leads to questionable accuracy of

microstructural parameters, which likely limits their sensitivity. In addition to the potential of diffusion parameters, relaxometry parameters such as T2 relaxation time have also shown capability to discriminate Gleason grades 3 and 4.(12, 15, 16) Works exploiting joint relaxation-diffusion analysis(20-23) (also in other organs, e.g. brain) have shown that these two types of parameters often contain complementary information.

In this work, we hypothesise that a unifying model capturing both relaxation and diffusion effects can enhance the accuracy of both parameters estimates and Gleason grade discrimination. We propose a new relaxation-VERDICT (rVERDICT) model to estimate jointly the diffusion and relaxation parameters in prostate. This work capitalizes on the VERDICT clinically feasible imaging protocol and exploits joint relaxation-diffusion analysis using machine learning. We demonstrate the benefits provided by rVERDICT comparing our results with the classic VERDICT and ADC from mp-MRI of the same patients.

MATERIALS AND METHODS

This study was performed with local ethics committee approval embedded within the INNOVATE clinical trial.(11) The trial is registered with ClinicalTrials.gov identifier NCT02689271.

Patient population and study design

72 men (median age, 64.8 years; range, 49.5–79.6 years) were recruited and provided informed written consent. The inclusion criteria were the following: (1) suspected PCa or (2) undergoing active surveillance for known PCa. Exclusion criteria included: (1) previous hormonal, radiation therapy or surgical treatment for PCa and (2) biopsy within 6 months prior to the scan. All patients underwent mp-MRI in line with international guidelines(24) on a 3T scanner (Achieva, Philips Healthcare, Best, Netherlands) supplemented by VERDICT DW-MRI (the clinical DCE part of mp-MRI was performed last after the VERDICT MRI).

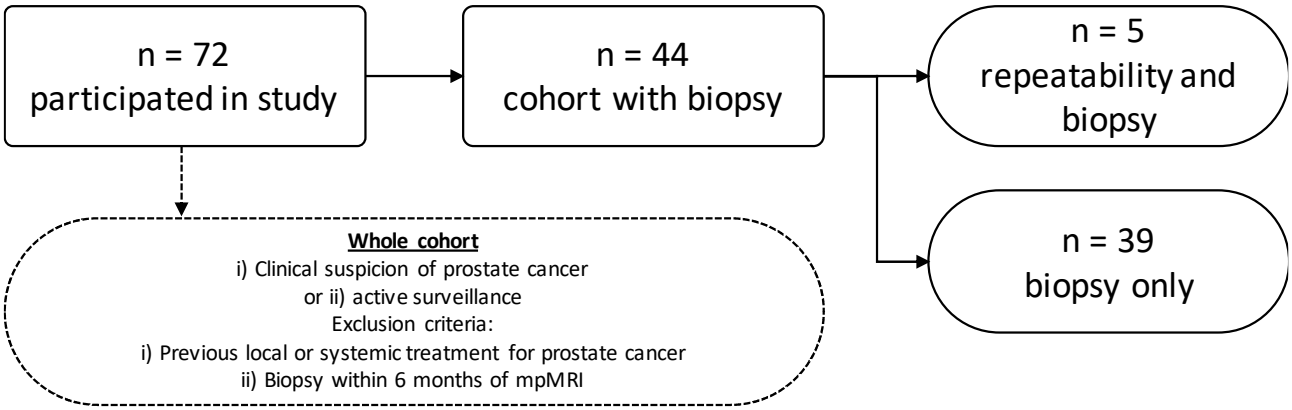

Figure 1 Participation flow diagram. mp-MRI = multiparametric MRI.

After the clinical mp-MRI and VERDICT DW-MRI, 44 participants underwent targeted transperineal template biopsy of their index lesion as clinically indicated. The mp-MRI was used to guide cognitive targeted template biopsy (performed by experienced urologists). Specialist genitourinary pathologists (A.F. and M.R) evaluated histological specimens stained with haematoxylin and eosin from the biopsy cores and assigned each biopsy core a Gleason grade. A participation flow diagram is presented in Figure 1.

MRI data acquisition

DW-MRI. The VERDICT protocol, adapted from(25), acquires DW-MRI data from the 44 "cohort with biopsy" (Figure 1) men using pulsed-gradient spin echo (PGSE) at five combinations (b, δ, Δ, TE, TR) of b-values b (in s/mm^2), gradient duration δ, separation Δ, echo time TE and repetition time TR (in ms): (90, 3.9, 23.8, 50, 2482); (500, 11.4, 31.3 65, 2482); (1500, 23.9, 43.8 90 2482); (2000, 14.4, 34.4, 71, 3945); (3000, 18.9, 38.8, 80, 3349), in three orthogonal directions using a cardiac coil. For each combination, a separate b=0 image was acquired. For b<100 s/mm^2 the number of averages (NAV) = 4 and for b > 100 s/mm^2 NAV= 6; voxel size 1.3×1.3×5 mm; matrix size 176×176; average signal-to-noise ratio (SNR) = 35; scan duration = 12 minutes 25 seconds.

Repeatability study. To evaluate the repeatability of VERDICT parameters, a scan-rescan repeatability study of the VERDICT DW-MRI acquisition protocol was performed in five participants. Participants were imaged twice, with less than 5 minutes between each examination (Figure 1).

Image Analysis

*DW-MRI data pre-processing***.** The pre-processing pipeline included denoising of the raw DW-MRI data using MP-PCA(26) as implemented within MrTrix3(27) 'dwidenoise', and then correction for Gibbs ringing(28) with custom code in MATLAB (The Mathworks Inc., Natick, Massachusetts, USA). To reduce possible artefacts caused by patient movement during scanning and eddy current distortions, we applied mutual-information rigid and affine registration using custom code in MATLAB.

Classic VERDICT. The VERDICT model is the sum of three parametric models, each describing the DW-MRI signal in a separate population of water from one of the three compartments: S_{ic} comes from intracellular water trapped inside cells (including epithelium), modelled as restricted diffusion in spheres of radius R and intra-sphere diffusivity D_{ic} = 2 μm^2/ms; S_{ees} comes from extracellular-extravascular water adjacent to, but outside cells and blood vessels (including stroma and lumen), modelled as Gaussian isotropic diffusion with effective diffusivity D_{ees} = 2 μm^2/ms; and S_{vasc} arises from water in blood undergoing microcirculation in the capillary network, modelled as randomly oriented sticks with intra-stick diffusivity D_{vasc} = 8 μm^2/ms. The total magnetic resonance signal for the multi-compartment VERDICT model is:

$$S(b)/S_0 = f_{vasc}S_{vasc}(D_{vasc}, b) + f_{ic}S_{ic}(D_{ic}, R, b) + f_{ees}S_{ees}(D_{ees}, b) \qquad [1]$$

where f_i is the proportion of signal from water molecules in population i = vasc, ic or ees, $f_{vasc} + f_{ic} + f_{ees} = 1$ and S_0 is the b=0 signal intensity. We refer to the original VERDICT work(5, 8) for further details on the specific expressions for S_{vasc}, S_{ic} and S_{ees}.

In classic VERDICT there are three free model parameters that are estimated by fitting equation [1] to DW-MRI data: f_{ees}, f_{ic}, R. The vascular signal fraction, f_{vasc}, is computed as 1 – f_{ic} – f_{ees}, while a cellularity index is computed as f_{ic}/R^3.

The VERDICT model for prostate is based on a few assumptions about water diffusion in the prostate tissue. First, the model assumes there are three major tissue compartments that mostly contribute to the measured DW-MRI signal: intra-cellular, intra-vascular and extra-cellular/extra-vascular, which are fully impermeable to water (i.e. no water exchange between compartments). Moreover, the water diffusivity inside the cellular and vascular compartments are fixed to D_{ic} = 2 μm^2/ms and D_{vasc} = 8 μm^2/ms. Several previous studies(5, 8, 9) investigated the validity of these assumptions under the experimental conditions of the optimized DW-MRI acquisition for VERDICT in prostate.(29)

One of the major limitations of the classic VERDICT formulation is that the signal fractions f_i in equation [1] are T2 and T1 relaxation weighted signal fractions. Therefore, their values and interpretation in terms of volume fractions of corresponding tissue compartments can be biased by the unaccounted relaxation properties of the tissue. This occurs when the DW-MRI acquisition includes different echo times (TE) and repetition times (TR) for different diffusion-weighting b, as in the imaging protocol for VERDICT DW-MRI.

Relaxation-VERDICT. The new relaxation-VERDICT (rVERDICT) model parameterises the T2 relaxation of the intracellular compartment by $T2_{ic}$, and that of vascular and extracellular/extravascular compartments by the same $T2_{vasc/ees}$. We also include the T1 relaxation contribution from the whole tissue as a single pool to account for the effect of different TR values used in the VERDICT DW-MRI acquisition. Mathematically, the rVERDICT model is

$$S(b,TE,TR) = S_0\left(1-e^{-\frac{TR}{T_1}}\right)\left[f_{vasc}^0 e^{-\frac{TE}{T2_{vasc/ees}}} S_{vasc}(D_{vasc},b) + f_{ic}^0 e^{-\frac{TE}{T2_{ic}}} S_{ic}(D_{ic},R,b) + f_{ees}^0 e^{-\frac{TE}{T2_{vasc/ees}}} S_{ees}(D_{ees},b)\right] \quad [2]$$

where we adopt the same terminology as classic VERDICT, but here the volume fraction parameters f^0_i, where i=vasc,ic,ees, avoid the bias in the corresponding classic VERDICT parameters from MR relaxation tissue properties.(16)

In total, rVERDICT has eight free parameters that are estimated by fitting equation [2] to DW-MRI data: S_0, T1, $T2_{ic}$, $T2_{vasc/ees}$, f^0_{ees}, f^0_{ic}, R and D_{ees}. For rVERDICT, the same assumptions as classic VERDICT for prostate apply, with the additional assumptions about the MR relaxation tissue properties as reported above and based on currently available experimental evidence.(17-19, 30) (17, 18, 20, 22, 23, 31, 32) Note that, unlike classic VERDICT, for rVERDICT we exploit the TE and TR dependence of the b=0 images for the T2 and T1 relaxation time estimation.

Model fitting with deep neural network (DNN). We obtained quantitative maps from both classic VERDICT and rVERDICT by fitting respectively equation [1] and [2] to the VERDICT DW-MRI data, using the signal averaged across the three gradient directions. To obtain ultra-fast parameters estimation, we performed the fitting using a DNN comprised of three fully connected layers,(33-35) trained on a large database of synthetic signals generated from each model with the addition of Rician noise corresponding to SNR = 35 (see Text S1 in Supplementary Material which reports further details).

DNN model fitting assessment. We tested the accuracy and precision of parameter estimates of rVERDICT using numerical simulations (see Text S2 in Supplementary Material which reports further details). To validate further the T2 relaxation time estimates, we compared the estimated $T2_{ic}$ and $T2_{vasc/ees}$ with independent T2 estimation from a T2-relaxometry MRI acquisition on seven patients, using the same DNN for fitting, trained with a two-compartment model of T2 relaxation (see Text S3 Supplementary Material which reports further details). We compared the distribution of estimated T2 values from rVERDICT with those from the T2-relaxometry acquisition for all the voxels within the prostate volume, and the corresponding mean and variance values. Statistically significant differences were assessed by two-tail t-test.

Regions-of-interest definition. Two board-certified experienced radiologists (reporting more than 2,000 prostate MR scans per year) manually placed ROIs on the VERDICT DW-MRI, guided by the standard mp-MRI.

Gleason grade classification

All ROI voxels were assigned to one of the corresponding five different histopathologic categories reported by the pathologists: benign, and Gleason grades 3+3, 3+4, 4+3 and ≥ 4+4.

Gleason classification with convolutional neural network (CNN). We estimated Gleason grade for each ROI with rVERDICT, classic VERDICT and ADC maps from the mp-MRI, using a CNN based on the SE-Res-Net(36) trained in a supervised fashion using all the ROIs and the associated five histopathologic categories (flowchart in Figure 2). To compensate for data imbalance, we used 5-fold cross-validation with stratified randomized folds to preserve the percentage of samples for each class (see Text S4 Supplementary Material which reports further details).

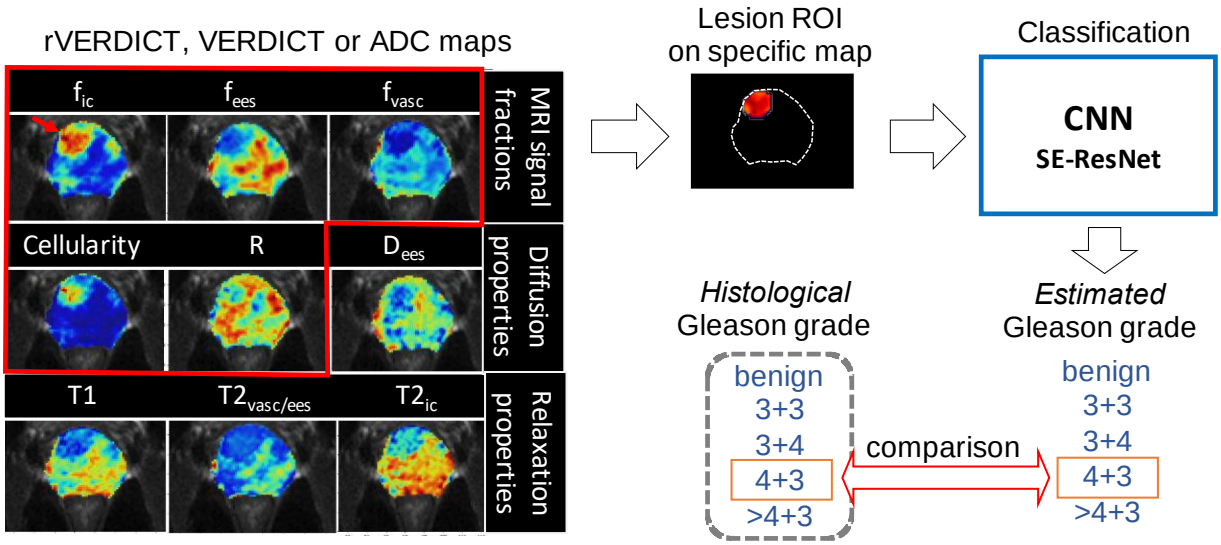

Figure 2 Flowchart of Gleason score classification using CNN. We classify the pre-specified lesion ROIs on the VERDICT maps using the SE-ResNet, whose input is an array the size of the full image but with zeros everywhere except within the lesion ROI (here we also show the boundaries of the prostate with dashed white line). The model then gives the corresponding Gleason score to the lesion minimizing the classification error with respect to the histological grading. Note that classic VERDICT can estimate only the parametric maps highlighted by the red contour.

Statistical Analysis

Scan-rescan reproducibility. We quantified repeatability using the adjusted coefficient of determination R^2 between each estimated model parameter in the first scan with the estimates from the second scan, considering the mean values of all the voxels within each ROI. We used subject-specific ROI instead of whole prostate statistics to remove potential bias due to deformation and different position of the prostate between the two scans. We used Bland-Altman plots and computed the coefficient of variation (CV) and the intraclass correlation coefficient (ICC).

Gleason grade classification performance. We quantified the performance of the 5-class CNN classification by computing the accuracy, precision and sensitivity (recall). From these, we computed summary statistics, such as the harmonic mean of precision and recall (F1-score) and the Cohen's kappa coefficient (kappa).(37)

RESULTS

Table 1 Summary of Demographic Data. Note: except where indicated, data are numbers of participants. Numbers in parentheses are ranges. PSA = prostate specific antigen.

Parameter	***Cohort with biopsy***
Number of Participants	44
Median age (y)	67 (49-79)
Median PSA level (ng/ml)	7.96 (0.83-72.11)
Highest Gleason grade of biopsied index lesion	
Benign	22
3+3	6
3+4	18
≥4+3	13
Median no. of total cores	23 (6-36)
Median no. of sites	2 (1-13)
Median no. of positive cores	5 (1-15)
Median maximum cancer core length (mm)	8 (1-14)
Median maximum cancer core length (%)	75 (10-100)
Median Prostate volume (ml)	43 (15-108)

There were 37 cancer lesions in the investigated cohort (n=44). Median prostate-specific antigen (PSA) level was 7.0 ng/mL (range, 1.0–71.0 ng/mL), the median time between VERDICT MRI and biopsy was 66.9 days (range, 8–167 days). Of the 37 cancer lesions, 6 were Gleason grade 3+3, 18 were Gleason grade 3+4, and 13 were Gleason grade 4+3 or greater. A summary of the demographic data is provided in Table 1.

Repeatability

The five participants of the scan-rescan repeatability study had median age 67.7 years (range, 50.0–82.0 years). The adjusted coefficient of determination R^2 for the diffusion and T2 relaxation parameters from rVERDICT ranged from 0.74 to 0.99; the coefficient of variation CV ranged from 1% to 10%; and the ICC from 78% to 98%. The correlation plots and Bland-Altman plots of all the rVERDICT parameters are reported in Figure S1 in Supplementary Material.

Numerical simulations showed that the DNN had similar accuracy but higher precision (as shown in Figure S2 in Supplementary Material) and robustness (as shown in FigureS3 in Supplementary Material) for all the rVERDICT parameters, compared to conventional non-linear least squares minimisation.

We found good agreement between the rVERDICT estimated distribution of T2 values in the prostate tissue and the T2 estimates from an independent multi-TE acquisition, with average values not statistically different (P>0.05). The estimated compartmental T2 values are also in agreement with current literature, with $T2_{ic}$~60 ms and $T2_{vasc/ees}$~250 ms.(17, 18, 20, 22, 23, 31, 32) The Figure S4 in Supplementary Material shows the comparison of rVERDICT T2 estimates and the multi-TE acquisition estimates for the best and worst performing cases.

Gleason grade classification

Table 2 Overall classification performances. Accuracy, precision, recall, F1-score and kappa for the apparent diffusion coefficient (ADC) maps obtained from multi-parametric MRI (mp-MRI), classic VERDICT and rVERDICT (using only the intracellular volume fraction fic maps or all the parametric maps).

Input	Accuracy	Precision	Recall	F1-score	Kappa
ADC mp-MRI	0.739	0.794	0.738	0.734	0.549
f_{ic} classic VERDICT	0.769	0.805	0.769	0.762	0.661
f^0_{ic} rVERDICT	0.846	0.865	0.846	0.838	0.773
All maps classic VERDICT	0.785	0.829	0.784	0.788	0.788
All maps rVERDICT	0.862	0.885	0.862	0.859	0.815

We obtained the best performances using the 5-class CNN classifier trained with all the rVERDICT parametric maps, achieving accuracy=0.862, F1-score=0.859 and kappa=0.815. For comparison, using the ADC map from mp-MRI we achieved accuracy=0.739, F1-score=0.734 and kappa=0.549, while using all the maps from classic VERDICT we achieved accuracy=0.785, F1-score=0.788 and kappa=0.788 (Table 2). The CNN trained and tested only with the intracellular volume fraction f^0_{ic} map from rVERDICT still outperformed both the CNN trained with f_{ic} from classic VERDICT and the ADC from mp-MRI (Table 2).

For per-class Gleason grade classification using all maps from rVERDICT, we found F1-score=0.857, 0.750, 0.889, 0.842 and 0.750 for the classes corresponding to benign, Gleason grades 3+3, 3+4, 4+3 and >4+3 respectively. All the performances on each Gleason grade classification, using rVERDICT, classic VERDICT and the ADC from mp-MRI are reported in Table 3, together with the corresponding precision and sensitivity (recall).

Three exemplar cases are shown in Figure 3 to demonstrate lesions with Gleason score 3+3 (Figure3a, green arrow), 3+4 (Figure3b, yellow arrow) and 4+3 (Figure3c, red arrow) on the DWI at b=2000 s/mm^2, ADC, and rVERDICT f^0_{ic}, D_{ees}, $T2_{vasc/ees}$ and T1 maps. While ADC does not correlate with Gleason grade in these cases, f^0_{ic} increases with Gleason grade, as also previously reported for classic VERDICT,(10) and D_{ees}, $T2_{vasc/ees}$ and T1 have lower values in the PCa lesions compared to benign tissue.

Table 3 Classification report (precision, recall, and F1-score) for each Gleason grade when using only the apparent diffusion coefficient (ADC) map from multi-parametric MRI (mp-MRI); all the parametric maps from classic VERDICT and all the parametric maps from rVERDICT with 5-class CNN (SE-ResNet).

Input	Class	Precision	Recall	F1-score
ADC mp-MRI	Benign	0.613	0.950	0.745
	3+3	0.857	0.600	0.706
	3+4	0.789	0.750	0.769
	4+3	1.000	0.600	0.750
	>4+3	1.000	0.400	0.571
All maps classic VERDICT	Benign	0.633	0.950	0.760
	3+3	1.000	0.800	0.889
	3+4	0.882	0.750	0.810
	4+3	0.857	0.600	0.706
	>4+3	1.000	0.600	0.750
All maps rVERDICT	Benign	0.818	0.900	0.857
	3+3	1.000	0.600	0.750
	3+4	0.800	1.000	0.889
	4+3	0.889	0.800	0.842
	>4+3	1.000	0.600	0.750

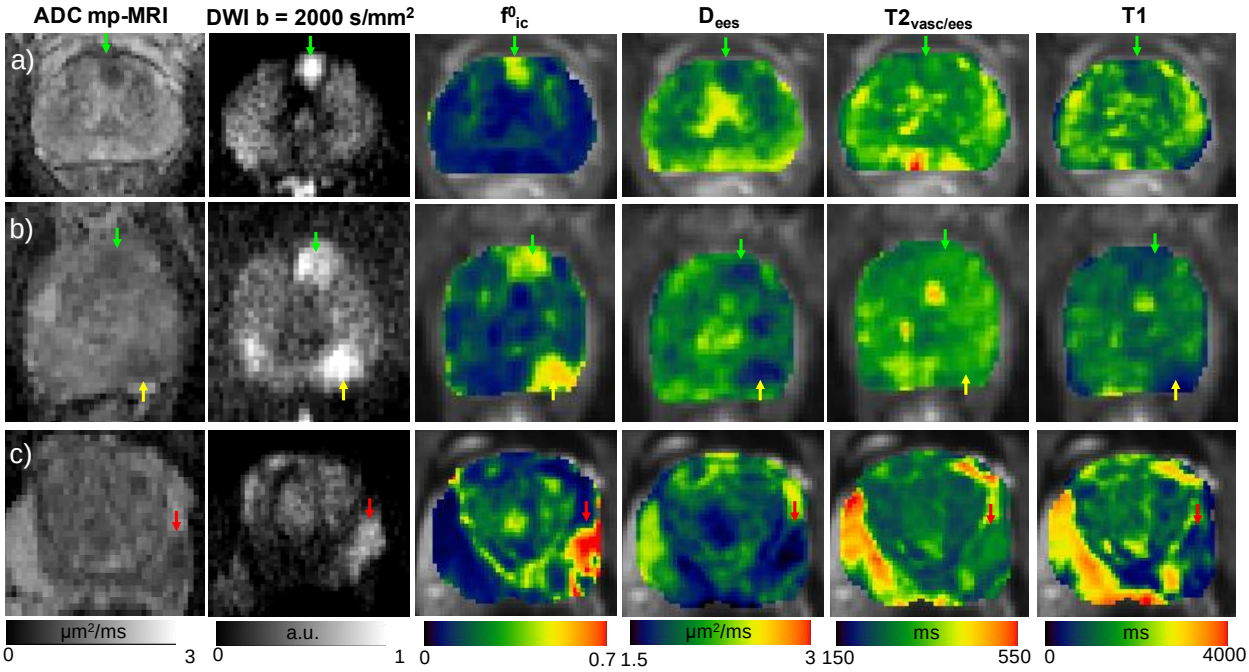

Figure 3 a-c) Apparent diffusion coefficient (ADC) maps from multi-parametric MRI (mp-MRI); diffusion-weighted image (DWI) at b = 2000 s/mm2 and rVERDICT maps (intracellular volume fraction f0ic; extracellular-extravascular apparent diffusion coefficient Dees; vascular/extracellular-extravascular T2 relaxation time T2vasc/ees; T1 relaxation time) for three exemplar patients with different PCa: a) age 58, PSA 4.78 Gleason 3+3 MCCL 7 mm; lesion in the anterior gland; b): age 67, PSA 5.21 left posterior lesion Gleason 3+4 MCCL 14 mm, left anterior lesion Gleason 3+3; c): age 68, PSA 8.68 Gleason 4+3, MCCL 10 mm; lesion in the left peripheral zone. Green arrows indicate Gleason grade 3+3, yellow arrows Gleason grade 3+4 and red arrows Gleason grade 4+3.

DISCUSSION

In this work, we propose a new prostate model called relaxation-VERDICT (rVERDICT) that provides joint estimation of relaxation and diffusion parameters, such as the intracellular T2 relaxation time ($T2_{ic}$) and the intracellular volume fraction (f_{ic}). Our hypothesis is that a unifying model capturing both relaxation and diffusion effects would enhance the accuracy of model parameters estimation and consequently the Gleason grade discrimination. As prostate histological components differ between Gleason grades,(38) we expect diffusion parameters, and in particular the f_{ic} from classic VERDICT that correlates strongly with epithelial volume fraction,(9, 39) to provide high biologic specificity to Gleason grade. However, classic VERDICT can only achieve discrimination of Gleason 3+3 from 3+4.(10) Gleason discrimination for higher scores like 3+4 from ≥4+3 is also important, as 4+3 cancers are associated with a three-fold increase in lethal PCa compared to 3+4 cancers.(40) Here we hypothesise that rVERDICT can compensate for any relaxation-induced bias that may be reducing the accuracy of classic VERDICT estimates, enabling more robust identification and discrimination of Gleason 4 components.

Our main results (Table 2 and Table 3) support our hypotheses, showing that the new information obtained from rVERDICT enables discrimination of five Gleason grades, with accuracy and Cohen's kappa 8 and 3 percentage points higher than classic VERDICT, and 12 and 24 percentage points higher than the ADC from mp-MRI, respectively. Most importantly, on the previously unattainable classification of Gleason grade 3+4 and 4+3, rVERDICT achieved the highest precision and sensitivity (both ≥80%) compared to classic VERDICT and ADC from mp-MRI.

The clinical utility of rVERDICT is demonstrated by the repeatability and fitting performance results. The f_{ic} from rVERDICT (f^0_{ic}) achieved higher repeatability (R^2 = 0.99; CV = 5%) compared to classic VERDICT f_{ic} (R^2 = 0.83; CV = 27%, as reported in(25)), showing that we can achieve greater reproducibility by removing confounds through disentangling relaxation from diffusion parameters. Additionally, the fitting approach based on deep neural network provides precise and robust estimates for all the rVERDICT parameters with dramatic reduction of the processing time (~35

seconds compared to ~15 minutes using the AMICO framework(25, 41)), enabling on-the-fly rVERDICT map generation. This is a critical point that further enables clinical translation of the technique.

A great advantage of rVERDICT is the possibility to obtain simultaneously diffusion and relaxation properties of prostate tissue using only a 12-minute DW-MRI acquisition. We showed that estimates of T2 and T1 relaxation times with rVERDICT match those from independent measurements and literature. The T1 relaxation estimates had lower values within tumour lesions than in benign tissue (mean 1576 vs 2754 ms, respectively, P=0.003), in line with estimates obtained using independent T1 measurements.(17-19, 30) The T2 relaxation estimates $T2_{ic}$ within tumours were similar for benign tissue (mean 61 vs 67 ms, respectively, P>0.05) and $T2_{vasc/ees}$ within tumour lesions were lower than in benign tissue (mean 300 vs 383 ms, respectively, P=0.023), in agreement with literature.(17, 18, 20, 22, 23, 31, 32) For seven patients in our cohort, we performed an independent multi-TE acquisition which allowed for a direct comparison of the estimated T2 values using the two methods. Results showed that the rVERDICT T2 values were not statistically different (P>0.05) from the independent T2 measurements.

Although multi-TE acquisitions have the advantage of offering MR images with higher resolution and less artefacts compared to DW-MRI, diffusion-based techniques give us unique insight into microstructure. rVERDICT harnesses the information from both relaxation and diffusion to provide a holistic characterisation of the prostate. In addition to the intracellular volume fraction f^0_{ic}, here we also showed that other rVERDICT parameters such as cellularity, f^0_{ees}, D_{ees} and T1 can offer complementary information to aid Gleason grade discrimination.

To demonstrate the potential of rVERDICT for improving PCa diagnosis, we present three example cases in Figure 3. The new information provided by rVERDICT maps can help improve the ability of identifying and distinguishing Gleason grades of PCa lesions in cases where both the ADC map and the high b-value DWI already show clear contrast (Figure 3a) and most importantly when these conventional measures provide ambiguous information (Figure3b and 3c).

There are several opportunities for further improvement in future work. The analysis presented here was performed on retrospective data with an acquisition protocol optimised for classic VERDICT probing a limited range of TE and TR values. This mostly compromises the sensitivity to long T2 values ($T2_{vasc/ees}$) and the reproducibility of measured T1 values. However, we have shown that our estimates are still in agreement with independent T2 and T1 measurements that cover a wider acquisition parameter space. Future work will explore optimization of the VERDICT MRI acquisition to explicitly account for T1 and long T2 relaxation times. Additionally, here we analysed only 44 patients for whom the biopsy results were available. This resulted in having limited and unbalanced different Gleason grades, hampering the possibility to examine differences in diagnostic performance (e.g. with comparison of areas under the receiver operating characteristic curve). However, we were still able to draw significant correlations and demonstrate the potential of rVERDICT. Further study will include rVERDICT analysis on larger cohorts. Finally, rVERDICT (as classic VERDICT) does not account for the effect of exchange and diffusion time dependence of water diffusivity,(20, 21) such as D_{ees}. However, the diffusion time used in this study was between 22-36 ms, a range for which previous studies have shown negligible effects due to permeability and time dependence.(20, 21, 42) Future work can explore the possibility to include these effects in the model and potentially estimate other tissue properties such as cell membrane permeability.

In conclusion rVERDICT together with machine learning allows for ultra-fast and robust microstructural estimation of diffusion and relaxation properties of prostate cancer. This enables classification of Gleason grades, potentially allowing the utilisation of rVERDICT for screening and avoidance of unnecessary biopsies and treatments.

ACKNOWLEDGMENTS

This work was supported by EP/N021967/1, EP/R006032/1 and by Prostate Cancer UK: Targeted Call 2014: Translational Research St.2, project reference PG14-018-TR2. M.P. is supported by the UKRI Future Leaders Fellowship MR/T020296/1.

REFEREENCES

1. Graham J, Kirkbride P, Cann K, et al. Prostate cancer: summary of updated NICE guidance. BMJ. 2014;348:f7524.
2. Lee SI, Hectors SJ**.** Prostate MRI: Toward Imaging Tumor Histology. Radiology. 2020;296(2):356-7.
3. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet. 2017;389(10071):815-22.
4. Dopfert J, Lemke A, Weidner A, Schad LR**.** Investigation of prostate cancer using diffusion-weighted intravoxel incoherent motion imaging. Magn Reson Imaging. 2011;29(8):1053-8.
5. Panagiotaki E, Chan RW, Dikaios N, et al. Microstructural Characterization of Normal and Malignant Human Prostate Tissue With Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumours Magnetic Resonance Imaging. Investigative Radiology. 2015;50(4):218-27.
6. White NS, McDonald C, Farid N, et al. Diffusion-weighted imaging in cancer: physical foundations and applications of restriction spectrum imaging. Cancer Res. 2014;74(17):4638-52.
7. Chatterjee A, Harmath C, Oto A**.** New prostate MRI techniques and sequences. Abdom Radiol (NY). 2020.
8. Panagiotaki E, Walker-Samuel S, Siow B, et al. Noninvasive Quantification of Solid Tumor Microstructure Using VERDICT MRI. Cancer Research. 2014;74(7):1902-12.
9. Bailey C, Bourne RM, Siow B, et al. VERDICT MRI validation in fresh and fixed prostate specimens using patient-specific moulds for histological and MR alignment. NMR Biomed. 2019;32(5):e4073.
10. Johnston EW, Bonet-Carne E, Ferizi U, et al. VERDICT MRI for Prostate Cancer: Intracellular Volume Fraction versus Apparent Diffusion Coefficient. Radiology. 2019;291(2):391-7.
11. Johnston E, Pye H, Bonet-Carne E, et al. INNOVATE: A prospective cohort study combining serum and urinary biomarkers with novel diffusion-weighted magnetic resonance imaging for the prediction and characterization of prostate cancer. BMC Cancer. 2016;16(1):816.
12. Sabouri S, Chang SD, Savdie R, et al. Luminal Water Imaging: A New MR Imaging T2 Mapping Technique for Prostate Cancer Diagnosis. Radiology. 2017;284(2):451-9.
13. Chan RW, Lau AZ, Detzler G, et al. Evaluating the accuracy of multicomponent T2 parameters for luminal water imaging of the prostate with acceleration using inner-volume 3D GRASE. Magn Reson Med. 2019;81(1):466-76.
14. Devine W, Giganti F, Johnston EW, et al. Simplified Luminal Water Imaging for the Detection of Prostate Cancer From Multiecho T2 MR Images. J Magn Reson Imaging. 2019;50(3):910-7.
15. Sabouri S, Chang SD, Goldenberg SL, et al. Comparing diagnostic accuracy of luminal water imaging with diffusion-weighted and dynamic contrast-enhanced MRI in prostate cancer: A quantitative MRI study. NMR Biomed. 2019;32(2):e4048.
16. Hectors SJ, Said D, Gnerre J, et al. Luminal Water Imaging: Comparison With Diffusion-Weighted Imaging (DWI) and PI-RADS for Characterization of Prostate Cancer Aggressiveness. J Magn Reson Imaging. 2020.
17. Kjaer L, Thomsen C, Iversen P, Henriksen O**.** In vivo estimation of relaxation processes in benign hyperplasia and carcinoma of the prostate gland by magnetic resonance imaging. Magn Reson Imaging. 1987;5(1):23-30.
18. Storaas T, Gjesdal KI**.** Assessment of biexponential T1 decay in prostate tissue. ISMRM 20th Scientific Meeting & Exhibition, 2012.
19. Baur ADJ, Hansen CM, Rogasch J, et al. Evaluation of T1 relaxation time in prostate cancer and benign prostate tissue using a Modified Look-Locker inversion recovery sequence. Sci Rep. 2020;10(1):3121.
20. Lemberskiy G, Fieremans E, Veraart J, et al. Characterization of prostate microstructure using water diffusion and NMR relaxation. Front Phys. 2018;6.
21. Lemberskiy G, Rosenkrantz AB, Veraart J, et al. Time-Dependent Diffusion in Prostate Cancer. Invest Radiol. 2017;52(7):405-11.

22. Chatterjee A, Bourne RM, Wang S, et al. Diagnosis of Prostate Cancer with Noninvasive Estimation of Prostate Tissue Composition by Using Hybrid Multidimensional MR Imaging: A Feasibility Study. Radiology. 2018;287(3):864-73.
23. Zhang Z, Wu HH, Priester A, et al. Prostate Microstructure in Prostate Cancer Using 3-T MRI with Diffusion-Relaxation Correlation Spectrum Imaging: Validation with Whole-Mount Digital Histopathology. Radiology. 2020:192330.
24. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. Eur Urol. 2016;69(1):16-40.
25. Bonet-Carne E, Johnston E, Daducci A, et al. VERDICT-AMICO: Ultrafast fitting algorithm for non-invasive prostate microstructure characterization. Nmr in Biomedicine. 2019;32(1).
26. Veraart J, Fieremans E, Novikov DS**.** Diffusion MRI noise mapping using random matrix theory. Magn Reson Med. 2016;76(5):1582-93.
27. Tournier JD, Smith R, Raffelt D, et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. Neuroimage. 2019;202:116137.
28. Kellner E, Dhital B, Kiselev VG, Reisert M**.** Gibbs-ringing artifact removal based on local subvoxel-shifts. Magn Reson Med. 2016;76(5):1574-81.
29. Panagiotaki E, Ianus A, Johnston E, et al. Optimised VERDICT MRI protocol for prostate cancer characterisation. Proc. Intl. Soc. Mag. Reson. Med. 23rd, 2015.
30. Yu AC, Badve C, Ponsky LE, et al. Development of a Combined MR Fingerprinting and Diffusion Examination for Prostate Cancer. Radiology. 2017;283(3):729-38.
31. Storas TH, Gjesdal KI, Gadmar OB, et al. Prostate magnetic resonance imaging: multiexponential T2 decay in prostate tissue. J Magn Reson Imaging. 2008;28(5):1166-72.
32. Grgac K, Li W, Huang A, et al. Transverse water relaxation in whole blood and erythrocytes at 3T, 7T, 9.4T, 11.7T and 16.4T; determination of intracellular hemoglobin and extracellular albumin relaxivities. Magn Reson Imaging. 2017;38:234-49.
33. Golkov V, Dosovitskiy A, Sperl JI, et al. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. IEEE Trans Med Imaging. 2016;35(5):1344-51.
34. Ye C, Li Y, Zeng X**.** An improved deep network for tissue microstructure estimation with uncertainty quantification. Med Image Anal. 2020;61:101650.
35. Barbieri S, Gurney-Champion OJ, Klaassen R, Thoeny HC**.** Deep learning how to fit an intravoxel incoherent motion model to diffusion-weighted MRI. Magnetic Resonance in Medicine. 2020;83(1):312-21.
36. Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks. IEEE Trans Pattern Anal Mach Intell. 2020;42(8):2011-23.
37. Cohen J**.** A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20(1):37-46.
38. Chatterjee A, Watson G, Myint E, et al. Changes in Epithelium, Stroma, and Lumen Space Correlate More Strongly with Gleason Pattern and Are Stronger Predictors of Prostate ADC Changes than Cellularity Metrics. Radiology. 2015;277(3):751-62.
39. Bonet-Carne E, Tariq M, Pye H, et al. Histological Validation of in-vivo VERDICT MRI for Prostate using 3D Personalised Moulds. Proc. Intl. Soc. Mag. Reson. Med., 2018.
40. Stark JR, Perner S, Stampfer MJ, et al. Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3? J Clin Oncol. 2009;27(21):3459-64.
41. Daducci A, Canales-Rodriguez EJ, Zhang H, et al. Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. Neuroimage. 2015;105:32-44.
42. Reynaud O**.** Time-dependent diffusion MRI in cancer: tissue modeling and applications. Frontiers in Physics. 2017;5:58.
43. Kingma DP, Ba J**.** Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

Supplementary Material

Text S1.
RVERDICT model fitting with deep learning

We obtain quantitative rVERDICT maps by fitting equation [2] to the DW-MRI data acquired at different b values, TE and TR. Note that for classic VERDICT the signal is normalized by the b = 0 of each b shell, while in this work we do not analyse the normalized signal, but we rather exploit the TE and TR dependence of the b = 0 in each shell for improving the T2 and T1 relaxation time estimation.

Deep learning fitting for parameter estimation. To obtain ultra-fast and robust parameter estimation, we perform the fitting using a deep learning approach based on deep neural networks (DNN),(33-35) using the 'MLPregressor' implemented in Python scikit-learn 0.23 (https://scikit-learn.org/stable/). The input of our DNN is a vector of DW-MRI signals for each b, TE and TR combinations (a total of 10 in this specific case). The DNN consists of three fully connected hidden layers with 150 neurons, each characterised by a linear matrix operation followed by element-wise rectified linear unit function (ReLU), and a final regression layer with the number of output neurons equal to the number of tissue parameters to be estimated (i.e. eight for rVERDICT). The DNN is optimised by backpropagating the mean squared error (MSE) between ground truth model parameters and DNN predictions. We perform the optimisation with the adaptive moment estimation (ADAM) method(43) for 1000 epochs (adaptive learning rate with initial value of 0.001; one update per mini-batch of 100 voxels; early stopping to mitigate overfitting; and momentum = 0.9) on 100,000 synthetic DW-MRI signals (split into 80% for training and 20% for validation). We generate the synthetic DW-MRI signals using equation [2] with different values for the model parameters randomly chosen between biophysical plausible intervals: S_0 = [0, maximum b=0 intensity x 2], T1 = [10, 4000] ms, $T2_{ic}$ = [1, 150] ms, $T2_{vasc/ees}$ = [150, 800] ms, f_{ees} and f^0_{ees} = [0.01, 0.99], f_{ic} and f^0_{ic} = [0.01, 0.99], R = [0.01, 15] μm and D_{ees} = [0.5, 3] μm^2/ms. We also add Rician noise corresponding to SNR = 35 to consider experimental noise effect. For the final parameter computation, we use the DNN at the epoch with minimum validation loss.

The creation of the training set and training of the DNN (to be done only once) took ~100 seconds (4 threads on Intel Core i7 processor at 2.4GHz). Prediction of the trained DNN for the whole unmasked DW-MRI dataset (~$5x10^5$ voxels) took ~35 seconds, for each subject.

Text S2.
Assessment of rVERDICT estimates accuracy and precision

To assess the accuracy and precision of the DNN estimator for a complex model as rVERDICT we perform numerical simulations with known ground-truth. Specifically, we generate 4^7 synthetic DW-MRI signals with equation [2] and all the possible combinations of the seven model parameters using four values linearly distributed in the same range as described in the previous section but fixing the S_0 = 1. First, we evaluate accuracy and precision in terms of bias and dispersion of the prediction compared to the known ground-truth. Then, for a given realistic choice of each of the seven model parameters: T1 = 2700 ms, $T2_{ic}$ = 70 ms, $T2_{vasc/ees}$ = 530 ms, f^0_{ees} = 0.40, f^0_{ic} = 0.40, R = 8 μm and D_{ees} = 2 μm^2/ms, we evaluate the impact of possible degeneracy and local minima by comparing the distribution of predicted model values for each of the chosen parameters with the known ground-truth, when the other six parameters were varied (i.e. 4^6 different combinations).

As benchmark, we compare the performance of our DNN model with the more conventional non-linear least squares minimisation, as implemented in the 'nonlincon' function in MATLAB, using a grid-search algorithm for the initial guess.

Text S3.
Assessment of rVERDICT estimates of multiple T2 relaxation times in prostate tissue

T2-relaxometry MRI acquisition. To additionally assess the performance of rVERDICT in estimating multiple T2 relaxation times in prostate, a multi-TE acquisition was also acquired for seven of these patients for an independent estimate of the multiple T2 relaxation times. The sequence consisted of a multi-echo spin-echo sequence with an echo spacing of 31.25 msec and TR = 8956 msec. The other imaging parameters were: number of echo times = 32; field of view (FOV) = 180×180 mm; acquired voxel size = 2×2×4 mm; scan duration = 5 minutes 50 seconds.

T2-relaxometry MRI data pre-processing. The pre-processing pipeline included only registration of the images at each TE to the image at the first TE, using a mutual-information rigid registration (six degrees of freedom) implemented in MATLAB.

T2-relaxometry MRI data analysis. To provide an independent T2 estimation of the relaxation times $T2_{ic}$ and $T2_{vasc/ees}$ assumed in rVERDICT, we fit the following equation to the multi-TE data:

$$M(TE) = A_{ic} \exp\left(-\frac{TE}{T2_{ic}}\right) + A_{vasc/ees} \exp\left(-\frac{TE}{T2_{vasc/ees}}\right) + \varepsilon \quad [3]$$

where A_i are the relative signal intensities of the compartment i = ic and vasc/ees, $T2_i$ are the corresponding T2 values, with $T2_{ic}<T2_{vasc/ees}$, and ε is a plateau constant, accounting for non-zero noise floor in magnitude images and found to improve significantly the T2 estimates.(31) For the fitting, we used the same DNN architecture described in Text, Supplemental Digital Content 1, trained on 100,000 synthetic signals with different random values between reasonable intervals: $A_{i=ic,vasc/ees}$ = [0, maximum b=0 intensity x 2], $T2_{ic}$ = [1, 150] ms, $T2_{vasc/ees}$ = [150, 800] ms, and ε = [0.01, 0.20]. We added Rician noise with SNR = 25 to match experimental noise conditions. We compare the distribution of estimated T2 values estimated from rVERDICT and the multi-TE acquisition for all the voxels within the prostate volume, as well as corresponding mean and variance.

Text S4.
Gleason grade classification through CNN

For the CNN model, we chose the SE-Res-Net(36) which gives the best performance with our data based on comparison with other networks architectures (ResNet, DenseNet). SE-ResNet combines ResNet with embedded “Squeeze-and-Excitation (SE)" block that adaptively recalibrates feature response (flowchart in Figure 2 of main text). SE blocks intrinsically introduce dynamics conditioned on the input, helping to boost feature discriminability. In SE-ResNet, SE was added before summation with the identity branch. Please refer to(36) for more details on the architecture.

For the data imbalance, we used 5-fold cross-validation (with test size 20%) with stratified randomized folds to preserve the percentage of samples for each class. For training, we crop the images selecting the ROIs’ bounding box to the same size 96x96 pixels and apply data augmentation using rigid transformation (affine and rotation). We train the network for 30 epochs with cross-entropy loss and Adam optimization (learning rate of 10^{-5}).

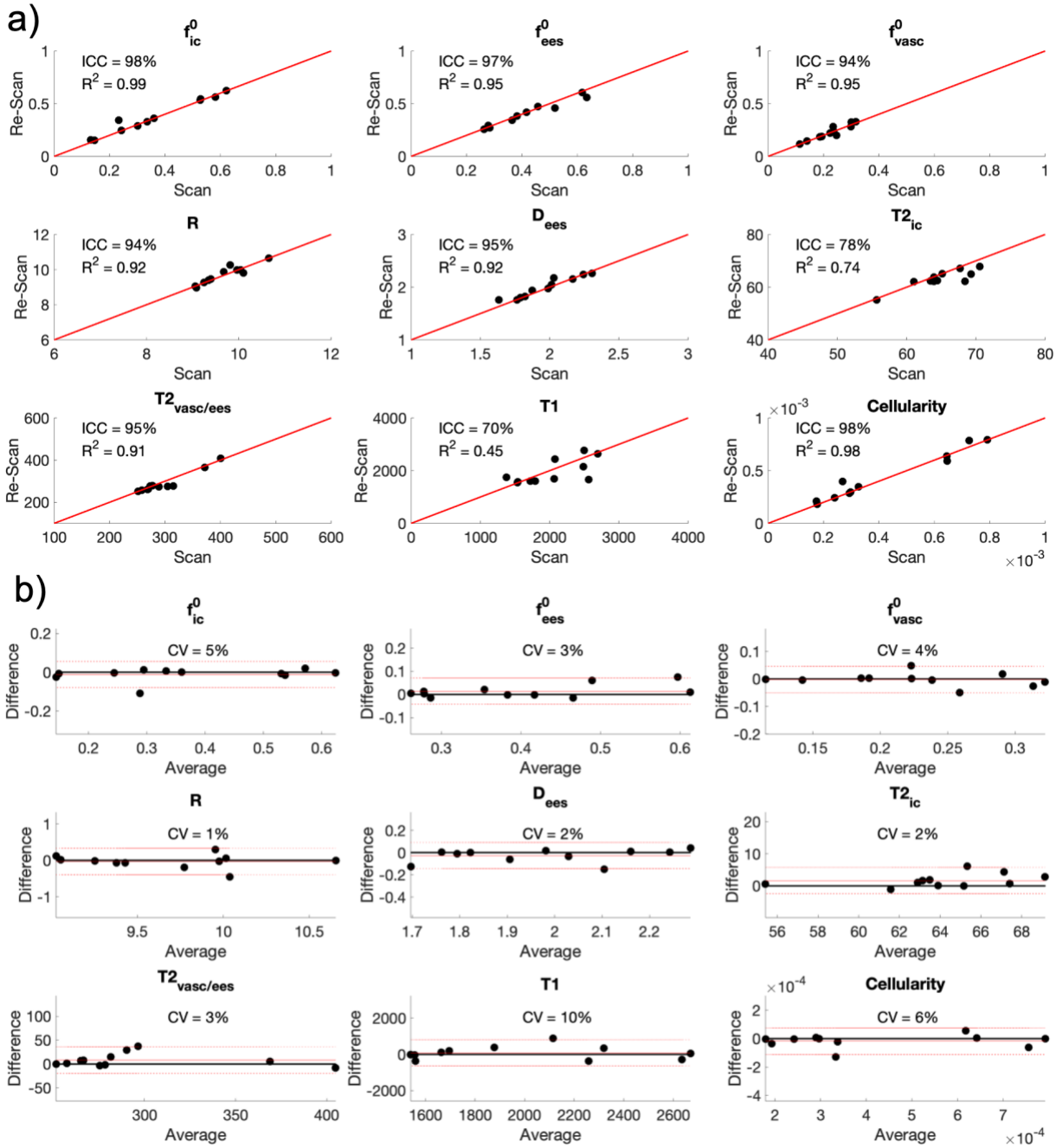

Figure S1. *Reproducibility of the rVERDICT parameters. a) Correlation plots for all the rVERDICT parameters in the scan/rescan study. The adjusted coefficient of determination R^2 and the intraclass correlation coefficient ICC are reported for each of them, together with the identity line. b) Bland-Altman plots of the difference between scan and re-scan versus the average for all the rVERDICT parameters. The coefficient of variation CV, computed as standard deviation over the mean, is reported for each of them, together with the average (straight red line) $\pm$ 1.96 standard deviation (dashed red lines) of the difference. The solid black line represents the ideal condition of zero difference. The dimensional parameters are in μm (apparent cell radius R); μm^{-3} (Cellularity); μm^2/ms (extracellular-extravascular apparent diffusivity Dees) and ms (T1, intracellular $T2_{ic}$ and vascular/extracellular-extravascular $T2_{vasc/ees}$).*

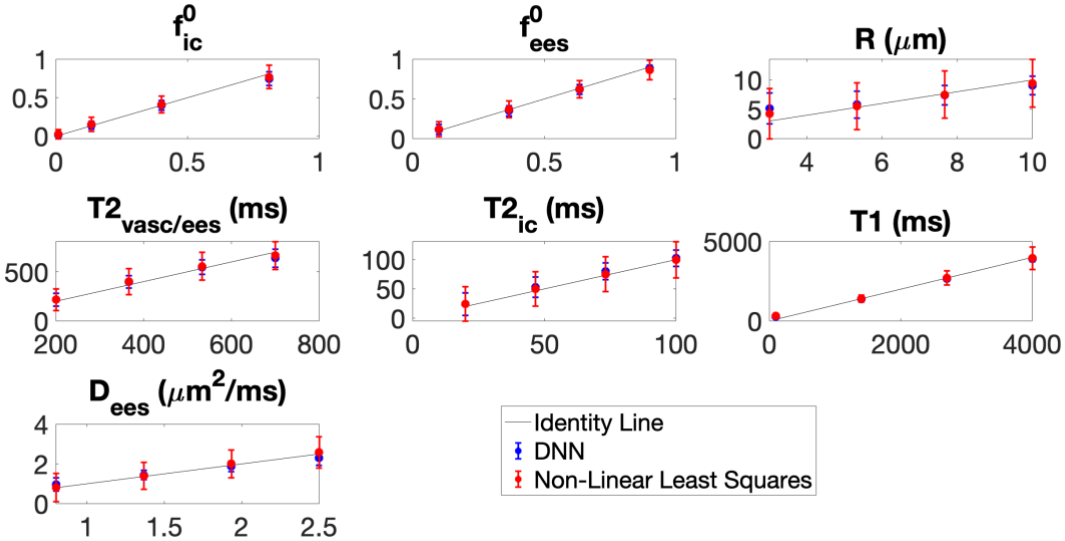

Figure S2. *Accuracy and precision of model fitting. The mean (data points) and variance (error bars) of the prediction for DNN and conventional non-linear least squares optimization are plotted against the known ground truth from numerical simulations. The identity line is also plotted to aid appreciating the accuracy of the prediction from both methods (higher the accuracy, closer the mean prediction to the identity line). The variance of the prediction (error bars) is a good indicator of the precision of the estimation: smaller the variance, higher the precision.*

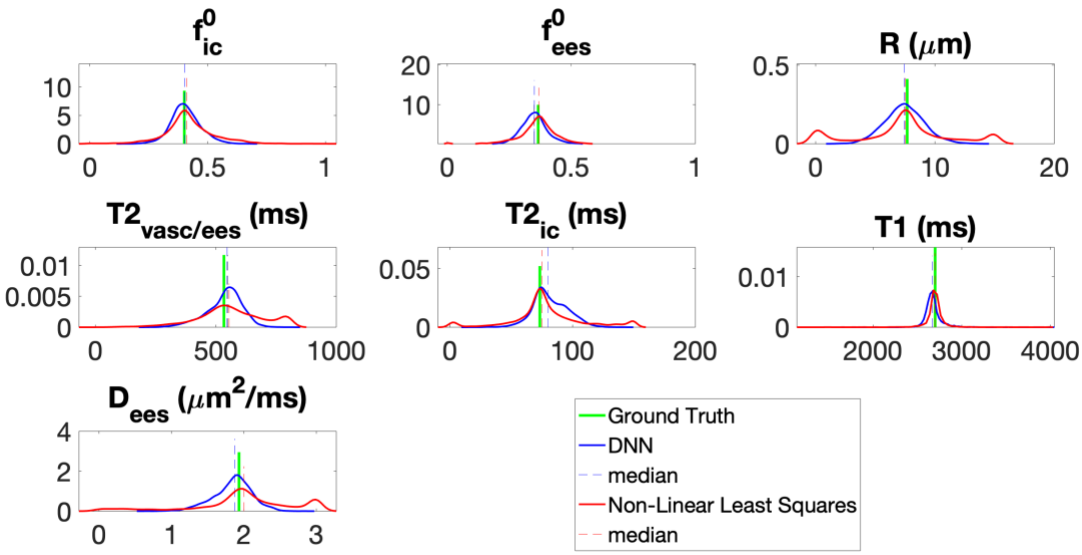

Figure S3. *Robustness of DNN estimation to degeneracy and spurious local minima. The probability density distribution of the estimates of the seven rVERDICT model parameters (S_0 was fixed to 1) are plotted for seven ground truth values (T1 = 2700 ms, $T2_{ic}$ = 70 ms, $T2_{vasc/ees}$ = 530 ms, f^0_{ees} = 0.40, f^0_{ic} = 0.40, R = 8 μm and D_{ees} = 1.9 μm^2/ms) and 4,096 different random combinations of the other parameters, for both DNN and conventional non-linear least squares optimisation. In absence of degeneracy and/or spurious local minima we would have very narrow distributions centred on the ground truth values. The wider the distribution, the less robust the estimation and the lower the precision due to degeneracy and/or spurious minima.*

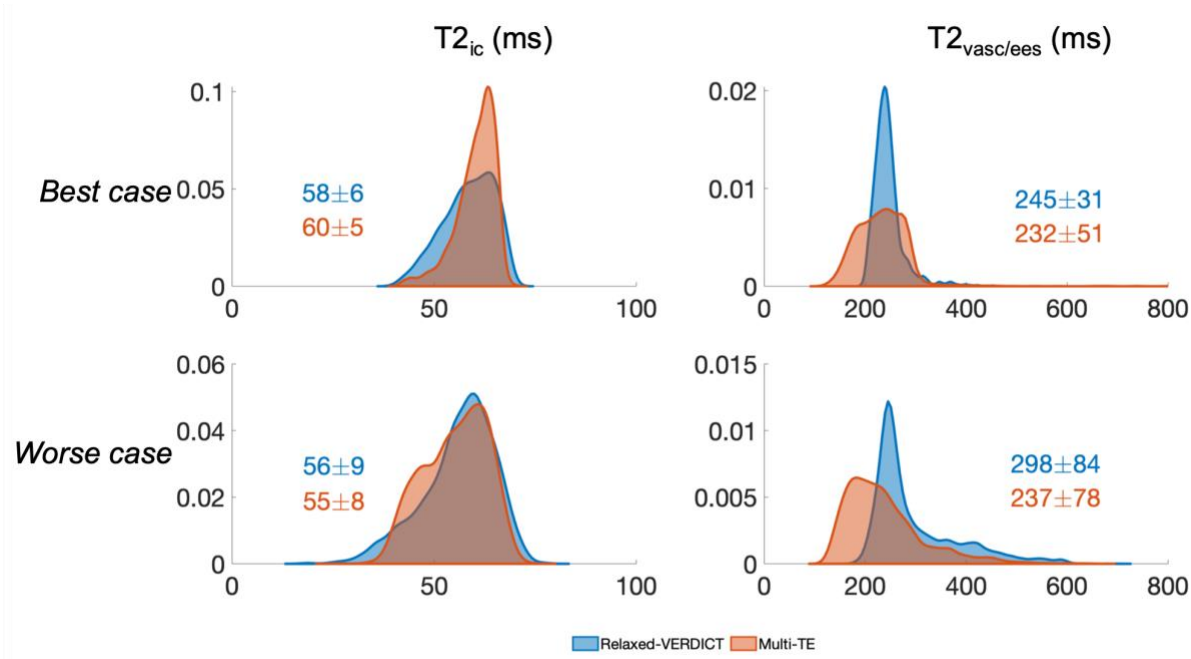

Figure S4. *Comparison of T2 estimates from rVERDICT (relaxed-VERDICT) and independent measurements using a multi-TE acquisition (multi-TE). The distributions show the probability density function of the estimated T2 values for all the voxels within the prostate volume. Mean and standard deviation values for each distribution are also reported.*