Aus der Klinik für Radiologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

# A 3D Multi-channel convolutional neural network for classification of prostate cancer using multiparametric MR imaging

# Ein 3D-Mehrkanal-Faltungsnetzwerk zur Klassifizierung von Prostatakrebs mithilfe der multiparametrischen MR-Bildgebung

zur Erlangung des akademischen Grades
PhD

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Nader Aldoj
aus Damaskus, Syrien

Datum der Promotion: 03.12.2021

# Table of contents

## List of Abbreviations

| | |
|---|---|
| ADC | Apparent diffusion coefficient |
| AUC | Area under the curve |
| CNN | Convolutional neural network |
| DCE | Dynamic contrast-enhanced |
| DWI | Diffusion-weighted imaging |
| MR | Magnetic resonance |
| Mp-MRI | Multiparametric magnetic resonance imaging |
| PCa | Prostate cancer |
| PI-RADS | Prostate imaging reporting and data system |
| PSA | Prostate-specific antigen |
| ROC | Receiver operating characteristics |
| TRUS | Transrectal ultrasound |
| T2w | T2-weighted |

# Abstract


## Abstract (English)

**Objectives:** To compare the diagnostic performance of the newly developed convolutional neural network (CNN) in classification of the significance of prostate cancer using multi-parametric magnetic resonance imaging (mp-MRI) against other similar approaches and the standard clinical assessment reported in the literature.

**Methods:** This is a retrospective study with a total of 200 patients (318 suspicious lesions) who received an MRI scan with three different pulse sequences (anatomic T2 weighted (T2w), diffusion-weighted imaging (DWI) with derivation of apparent diffusion coefficients (ADCs), and K-trans from dynamic contrast-enhanced MRI (DCE-MRI)), which served as an input for the CNN that is presented in this study. On the other hand, each patient in the dataset has one or more prostate lesions with their corresponding biopsy results which were used as the training and test labels. This study presents a novel neural network architecture that processes 3D images directly with the least manual interaction possible. This network was trained and evaluated on different individual and combinations of mp-MRI sequences which, in turn, outlines the diagnostic role, the benefits and the effects of each individual MRI sequence on the overall performance of the network. The obtained results were evaluated using widely used statistical values such as accuracy, area under the curve (AUC) sensitivity and specificity.

**Results:** When using the receiver operating characteristic (ROC) curve analysis, our 3D developed network had the highest average AUC value of 89.7% for input combination of ADC, DWI and K-trans while the rest of the mp-MRI combinations resulted in a significantly inferior performance in terms of AUC, sensitivity and specificity where p-value of 0.00025 was obtained when using T2w, ADC, and DWI; and 0.02 when using T2w and K-trans. Individual mp-MRI sequences had an AUC between 89.0% (88.6% sensitivity and 90.0% specificity) to 91.0% (81.2% sensitivity and 90.5% specificity). The effect of lesion size and volume was tested and showed no significant effect on the network's performance.

**Conclusion:** Our presented study shows that the 3D developed network which, requires minimal manual interactions, can process an input of mp-MRI and achieves very comparable results to the performance values of the reported experienced radiologist. Furthermore, this study shows that the size and the volume of the suspicious lesion have no significant effect on the performance of the network.

# Abstract


## Abstrakt (Deutsch)

**Zielsetzung:** Vergleich der diagnostischen Leistung des neu entwickelten convolutional neural network (CNN) bei der Klassifizierung der Bedeutung von Prostatakrebs mithilfe der multiparametrischen Magnetresonanztomographie (mp-MRT) mit anderen ähnlichen Ansätzen und der in der Literatur angegebenen klinischen Standardbewertung.

**Methoden:** Dies ist eine retrospektive Studie mit insgesamt 200 Patienten (318 verdächtige Läsionen), die einen MRT-Scan mit drei verschiedenen Pulssequenzen (anatomische T2-gewichtete (T2w), diffusionsgewichtete Bildgebung (DWI) mit Ableitung scheinbarer Diffusionskoeffizienten (ADCs) erhielten. und K-trans aus der dynamischen kontrastmittelverstärkten MRT (DCE-MRT), die als Eingabe für das in dieser Studie vorgestellte CNN diente. Andererseits weist jeder Patient im Datensatz eine oder mehrere Prostata-Läsionen mit den entsprechenden Biopsieergebnissen auf, die als Trainings- und Testetiketten verwendet wurden. Diese Studie präsentiert eine neuartige neuronale Netzwerkarchitektur, die 3D-Bilder direkt mit der geringstmöglichen manuellen Interaktion verarbeitet. Dieses Netzwerk wurde an verschiedenen Individuen und Kombinationen von mp-MRI-Sequenzen trainiert und bewertet, was wiederum die diagnostische Rolle, die Vorteile und die Auswirkungen jeder einzelnen MRI-Sequenz auf die Gesamtleistung des Netzwerks umreißt. Die erhaltenen Ergebnisse wurden unter Verwendung weit verbreiteter statistischer Werte wie Genauigkeit, Empfindlichkeit und Spezifität der Fläche unter der Kurve (englisch: area under the curve, kurz: AUC) bewertet.

**Ergebnisse:** Bei Verwendung der ROC-Kurvenanalyse (englisch: Receiver Operating Characteristic) hatte unser 3D-entwickeltes Netzwerk den höchsten durchschnittlichen AUC-Wert von 89,7% für die Eingangskombination von ADC, DWI und K-trans, während der Rest der mp-MRI-Kombinationen zu einem signifikant schlechteren Ergebnis führte Leistung in Bezug auf AUC, Empfindlichkeit und Spezifität, wobei ein p-Wert von 0,00025 erhalten wurde, wenn T2w, ADC und DWI verwendet wurden; und 0,02 bei Verwendung von T2w und K-trans. Einzelne mp-MRI-Sequenzen hatten eine AUC zwischen 89,0% (88,6% Sensitivität und 90,0% Spezifität) bis 91,0% (81,2% Sensitivität und 90,5% Spezifität). Der Effekt von Läsionsgröße und -volumen wurde getestet und zeigte keinen signifikanten Effekt auf die Netzwerkleistung.

**Schlussfolgerung:** Unsere vorgestellte Studie zeigt, dass das in 3D entwickelte Netzwerk, das nur minimale manuelle Interaktionen erfordert, eine Eingabe von mp-MRI verarbeiten kann und sehr vergleichbare Ergebnisse mit den Leistungswerten des berichteten erfahrenen Radiologen erzielt. Darüber hinaus zeigt diese Studie, dass die Größe und das Volumen der verdächtigen Läsion keinen signifikanten Einfluss auf die Leistung des Netzwerks haben.

# Synopsis

## 1. Introduction

### 1.1. background

The integral role of mp-MRI (which combines anatomic T2w with functional sequences) using the Prostate Imaging Reporting and Data System (PI-RADS) in prostate cancer (PCa) diagnosis continues to be developed and catches more attention over the past few years [1], [2] as PCa is the second leading cause of cancer death and the most common type of cancer in men [3]. Prostate MRI interpretation is a very challenging task due to heterogeneity in the signal received from the benign prostatic hyperplasia (BPH), physiological changes, inflammation, and the after-biopsy scars which, in turn, have a similar appearance to PCa and can be misclassified as such [4]. Screening of PCa is usually done with Digital rectal examination (DRE) and the level of prostate specific antigen (PSA) in the blood. Unfortunately, PSA is highly sensitive but not specific to PCa as it can result in a high value for BPH and low value for PCa in some cases [5]. Such methods (DRE and PSA) are still commonly used in clinical practice. However they often contribute to overdiagnosis and overtreatment due to their false positive rate and poor specificity [6] [7].

Alternatively Transrectal ultrasound (TRUS)-guided biopsy is currently the standard test for assessment of the tumour aggressiveness. These biopsies are assessed histologically with Gleason scores [8]. In general, however, this examination is blind to the position of the lesion and there is also a risk of missing the lesion. This test in conjunction with mp-MRI is therefore recommended, which encourages the localization of suspected lesions and thus improves the diagnostic accuracy of TRUS [9].

Mp-MRI, which includes anatomical sequence such as T2w, and functional ones such as diffusion-weighted imaging, apparent diffusion coefficient mapping, and dynamic contrast-enhanced imaging, became recently the method of choice and gained more importance as a more accurate and non-invasive imaging modality for detection, localization and characterization of PCa [10] [11] [12]. Nonetheless, mp-MRI interpretation involves well-trained radiologists (with many years of experience) and does not always attain the necessary sensitivity of 0.91 and precision of 0.81 recorded by for a 10 years experienced radiologist [9] [13]. Computer-aided diagnosis (CAD) and assessment can correct these limitations, and can greatly improve and increase human efficiency in PCa identification and characterization. CAD and machine learning methods have previously been shown to allow PCa to be identified and characterized automatically in several studies. [14] [15]. These methods relied on the manual extraction of features of suspicious areas such as texture properties, tissue heterogeneity, measurements of sizes and volumes, or border's properties and then training a machine learning classifier such as support vector machine (SVM), gradient boost (G-boost) or logistic regression, etc to determine whether or not the lesion under investigation is malignant, and to characterize lesion severity and tumour grade. The aforementioned methods rely usually

on hand engineered features which, in turn, require considerable amount of engineering skills and expertise in features extraction and selection in order for the classifier to benefit from the extracted information. However, the machine may not benefit much form those extracted features. In contrast, deep learning (DL)-based approaches extract and learn their own features during the training process which could yield the best performance possible, and result in better optimization and generalization [16]. Thus, recently, DL in general and convolutional neural networks (CNNs) in particular showed a substantial success over many difficult tasks, surpassing the performance of state-of-the-art methods that were among the best in many applications for years. CNNs became the method choice in many computer vision applications and showed an impressive generalizability when trained on large datasets [17] [18]. Despite the fact of minimal data, which typically poses a massive challenge that affects the generalizability of the algorithm, CNNs even had decent success in many medical applications and imaging tasks, not to mention that images obtained by diagnostic imaging modalities show broad variety due to the use of multiple types of scanners, procedures and standards of acquisition, making their generalization a very challenging task [19] [20]. Additionally, this poses a problem known as domain-shift which makes the trained model such as CNN fails at processing similar images obtained from a different scanner or different imaging sequence even if the imaging parameters are identical [21].

### 1.2. Related studies

Several apporaches have demonstrated the usefullness of CNNs in both the detection and characterization of PCa. Le et al [22] used multimodal input data (T2w and ADC images) on three well-known neural networks VGGNet [17], Inception [23], ResNet [24] and fused the resulted features with additional hand-engineered features. It demonstrated that in contrast to previously published findings, they could boost network efficiency by integrating the extracted features with hand-crafted features. Pre-training a network on imaging data and fine tune it on specific data has caught special attention in the last few year due to the improved accuracy and performance values that might be gained, Yang et al [25] investigated a co-trained fine-tuned version of an inception-like network [23], that was previously trained on real-life images, showing that fine-tuned version performed better than the one trained from scratch with random initializations of the learned parameters. Chen et al [26]used transfer learning on InceptionV3 and VGG-16 models that were originally trained on the ImageNet dataset, while Song et al [27] investigated a patch-based approach with 131 layers based on VGGNet. The AUCs achieved were 0.81, 0.83 and 0.944 for inceptionV3, respectively, and VGGNet from Chen et al and Song et al. On the other hand, Kiraly et al [28] suggested using a convolutional encoder-decoder that achieve both identification and severity classification of lesions and an average AUC of 0.834 was achieved. However, all studies listed so far All studies listed so far, however, have either used pre-delineated regions (manually segmented regions) and evaluated these areas, or rendered a collection of slices, which in turn involves careful selection of a representative slice to ensure optimum algorithm efficiency. Thus, 3D-based approach that can process 3D images directly with minimal or no interaction (such as lesion segmentation or slice selection) is of a great importance and can make the use

of the developed network more realistic and less prone to human error. Recently, Mehrtash et al [29] showed the viability of 3D CNNs to specifically interact with 3D images. They used a 3D cropped region of the images centred around the lesion under investigation as 3D input for the CNN and reached a comparable value of area under the curve (AUC) to that of a human radiologist using PI-RADS version 1 (v1) and version 2 (v2).

### 1.3. Our contribution

In our research, we intended to investigate the possibility of using deep learning-based PCa classification and characterization techniques as a step towards facilitating clinical workflow with the least amount of human interaction possible. We developed a 3D semi-automatic approach, which was based on the state-of-art convolutional neural networks, to predict the probability of a given prostate lesion to be clinically significant. This developed CNN can process multimodal 3D images directly without any need of lesion delineation or slice selection. The estimated center of the target lesion is the only necessary and manually determined parameter in our method, while all other processing steps are fully automated. We used a 3D multimodal cropped area around the alleged lesion as an input and transferred it into our CNN as an input, and the output of the CNN is the probability of a given lesion to be clinically significant.

Furthermore, the importance of different pulse sequences was studied to highlight the role of each of these sequences on the network's performance. Thus, we compared all possible variations of pulse sequences acquired as an input to our algorithm by mp-MRI to test how each of these influenced the network output. The four major classes are: group 1 where all MR sequences (T2w, ADC, DWI, and K-trans) have been used; group 2 where the T2w sequence has been eliminated. Just T2w and K-trans were in group 3, while group 4 consisted of T2w, ADC and DWI. However, additional groups and individual pulse sequences were tested and reported so that the impact of each of the sequences could be studied alone and together with each combination possible.

To assess the classification performance of our network and compare it to other reported studies, we performed receiver operating characteristics (ROC) curve analysis to calculate the area under the curve (AUC) [30], sensitivity and specificity for comparison with published results on the performance of experienced radiologists [13].

# 2. Methods

## 2.1. Problem formulation

The problem that we address in this study is a binary classification problem in its essence since the purpose of this research is to differentiate between PCa lesions that are clinically significant or nonsignificant. In detail, we use a 3D volume of individual or combined pulse sequences cropped around the centre of the suspicious lesion, and the output is a binary label of 0 or 1. Let us denote the input as $I$ and the output as $y$. The developed network, represented by equation (1), is characterised by weights and biases that are also known as learned parameters and can be denoted as Φ and β respectively.

$$y = F_{(\Phi,\ \beta)}(I) \qquad (1)$$

We can use equation (2) to measure the loss function for each input (individual or combination of sequences) in the dataset.

$$F(I, y) = -ylogp(Y = 1|I) - (1 - y)\log(Y = 0|I) \qquad (2)$$

where $p(y = i|I)$ denotes the probability of a certain output of the network.
The goal during the algorithm training is to minimize this loss function, in other words, minimize the error between the desired output and the actual ground truth label, so that the prediction of the algorithm is as close as possible to the actual ground truth.

## 2.2. Patient data

A well-organized dataset of mp-MRI for prostate cancer was published on the internet by SPIE-AAPM-NCI PROSTATEx challenge [31] [32] [33] to enable groups from across the globe to build methods for PCa diagnosis and classification. Initially, this dataset consists of 204 patients with their respective histological tests serving as the training labels, and 140 patients for testing where their labels are hidden from the participating groups.
As the purpose of our study here was not to participate in the challenge since the challenge was over when starting the project, an mp-MRI dataset of 200 patients with a total of 318 suspicious prostate lesions (patient can have one or multiple lesions) was used. Four patients were excluded due to their own relatively poor image quality. This dataset contains 243 clinically nonsignificant and 75 significant lesions and was divided into 175 patients for training and 25 patients for testing. We could not use the original test set (140 patients) released by the challenge organization because the ground truth labels were not publically provided.
The number of lesions is different for patients in the dataset; some patients have just one lesion, while others have two or more, see Figure 1 left side. The maximum number of lesions per patient in our dataset was 4 lesions which was present only in two patients. The right side

of Figure 1 shows the distribution of lesion according to their significance which indicates that the dataset is not balanced with respect to lesions classes.

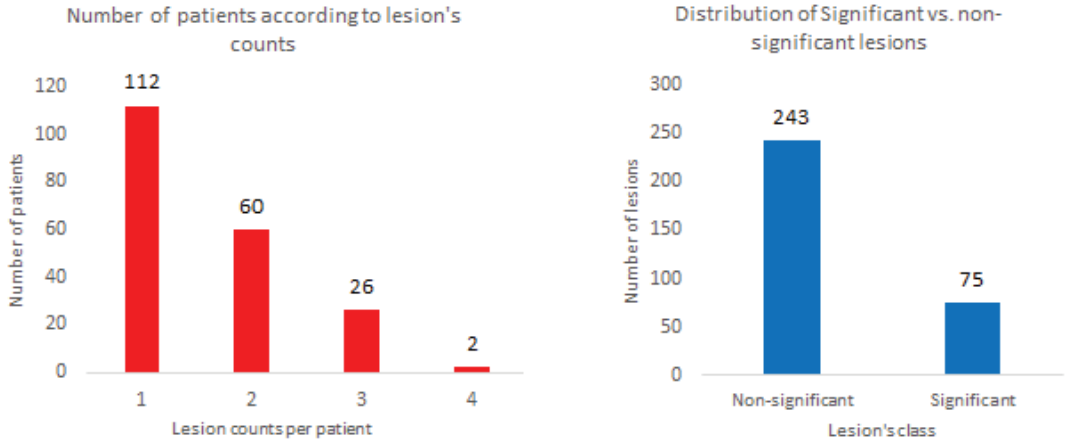


**Figure 1:** Analysis of lesions counts: Figure on the right side shows the number of significant vs. non-significant lesions. Figure on the left side shows number of patients with respect to lesions counts per patients.

The dataset contains information about the Gleason grade group (GGG), which is not used in this study since this study focuses on binary classification of lesions into significant vs. nonsignificant lesions. However, Figure 2 shows the number of lesions with respect to their GGG which, in turn, indicates that most of the suspicious lesions are of a grade 0 which are in fact normal prostatic tissues that are miss-classified as suspicious tissue due to their intrinsic heterogeneity.

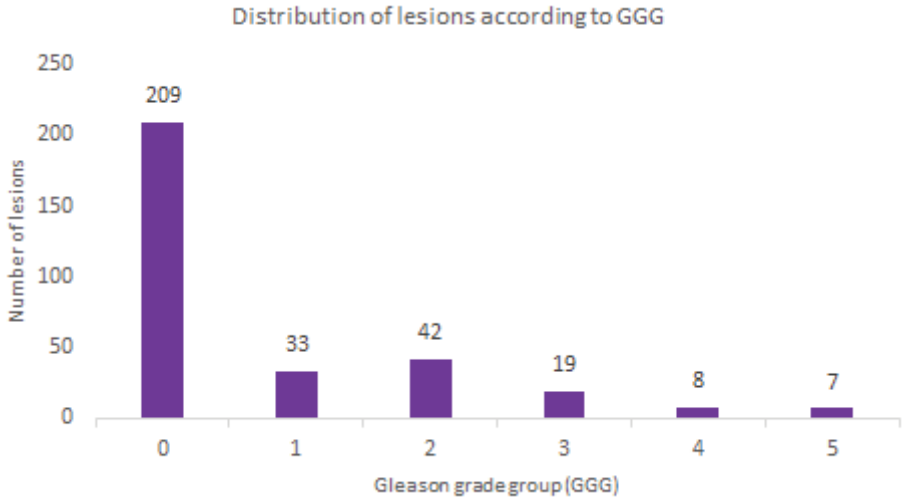


**Figure 2:** Distribution of lesions counts with respect to their Gleason grade group (GGG)

As neural networks are sensitive to class distribution and can be heavily biased towards the dominant class, we needed to ensure a similar distribution of lesions in the training set and the test set to avoid any performance bias. Thus, we performed zonal analysis to determine the zonal distribution of lesions within the prostate zones, see Figure 3. Most prostate lesions were located in the peripheral zone, followed by the transitional zone, the anterior stroma,

and the seminal vesicles. We split the images so that both training and test set have a similar distribution of lesions and the performance bias is reduced.

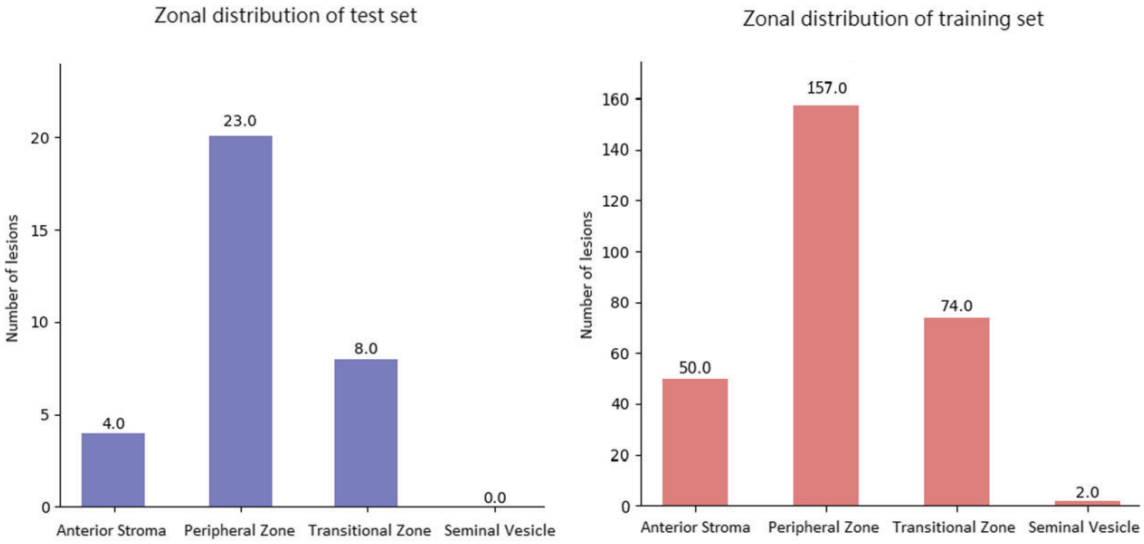


**Figure 3:** Distribution of prostate lesions by prostate zone in the test set and the training set. The bar graphs confirm a similar zonal distribution of lesions in both sets. This figure is from Aldoj et al. [34].

As we have mentioned earlier, this dataset is publically available and it comprises MR images obtained by mp-MRI, which consists of T2-weighted images, apparent diffusion coefficient maps, diffusion-weighted images, and K-trans images, which are obtained using dynamic contrast-enhanced MRI (DCE-MRI). The dataset contains biopsy-based histological findings for each lesion (clinically significant vs. nonsignificant lesion), which constituted the ground truth for training and testing. Moreover, the coordinates of each lesion in the image are provided so that the lesion centre can be easily located for any type of processing, delineation or analysis needed. The significance of the lesion was determined using Gleason scores where lesions with a Gleason score of 7 or above were classified as clinically significant lesions, those with lower Gleason scores as nonsignificant.

### 2.3. MR imaging protocol

The MR imaging protocol has been described before [34]. Briefly, all images in the dataset were acquired on 3T MAGNETOM Trio and Skyra MR Siemens scanners. T2-weighted images were acquired with a turbo spin echo sequence with 0.5 mm in-plane resolution and 3.6 mm of slice thickness. A single-shot planar imaging sequence was used for DWI with 2 mm in-plane resolution and the same slice thickness as for T2w. Diffusion-encoding gradients were used in three directions and b-values of 50, 400, and 800 were used to calculate ADC maps [31] [32] [33].

## 2.4. Image preprocessing

Before the mp-MR images could be used as input for the network, a preprocessing and preparation phase was needed. This was necessary because the images acquired with the different pulse sequences differ in voxel size and in-plane resolution. Thus to ensure the same resolution of all images in all sequences, resampling was required. This step was done using the image processing library (SimpleITK) to resample all images using bilinear interpolation to a voxel scale of 0.5, 0.5, and 3.0 mm in the x, y, and z directions respectively. Afterwards, all misaligned sequences were registered, using MITK workbench, with each other using manual rigid registration which was driven by six parameters (translation and rotation around x, y, and z axis), and then the images were cropped around the centre of the lesion which was provided as 3D coordinates in the original dataset. This copping process was done using a spherical cropping window with a radius of 20 mm. Since the cropping was not done in the voxel space, but rather relative to the scanner coordinate's system, the resulted volume differed slightly in dimensions from image to image. Thus, the resulted cropped volume was then contained in an empty predefined cube of a volume of 74, 74, and 14 voxels in x, y, and z direction, respectively, see Figure 4.

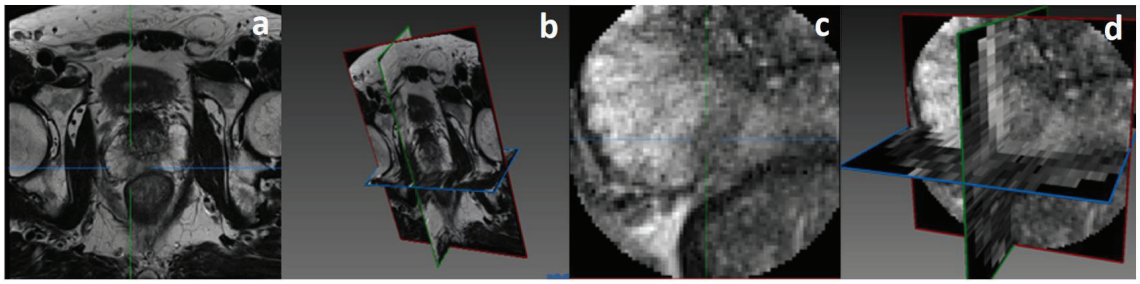


**Figure 4:** Spherical cropping volume: a and b show the original t2w image before cropping in 2D and 3D, respectively; c and d illustrate the cropped spherical region that is contained in the aforementioned predefined volume. This figure is from Aldoj et al. [34].

## 2.5. Network architecture

The PCa lesions in the dataset just outlined were classified using a very popular type of network known as convolutional neural network (CNN). We developed a 3D semiautomatic CNN-based approach to predict the probability of a given prostate lesion to be clinically significant. As an input, we used a 3D multimodal (e.g. ADC, K-trans, DWI, and T2w) cropped

regions around the suspected lesion and passed them as an input into our CNN where a final classification layer was represented by a softmax function, see Figure 5.

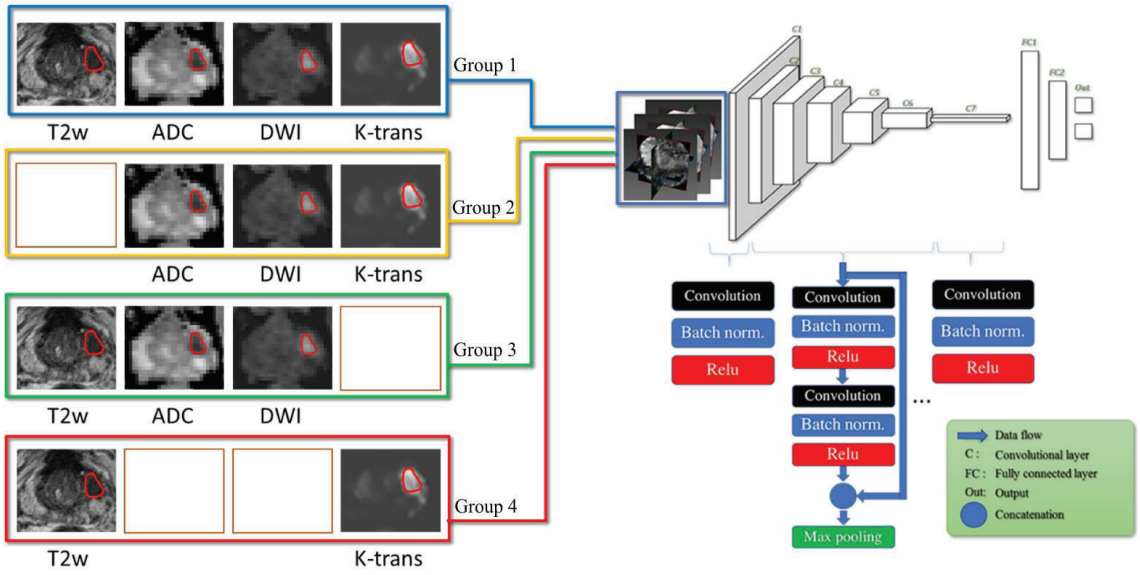


**Figure 5:** Illustration of the proposed approach: the figure shows the network architecture on the left and the four main input groups on the right. Part of this figure is from Aldoj et al. [34].

This network consists of different stages of convolutional layers, followed by two densely connected layers and a final softmax layer with two outputs that encode the actual probability of the predictions. Each of the convolutional stages in the graph has at least one combination of convolutional layers followed by a batch normalization, rectified linear unit (RelU) activation and a max-pooling layer. All convolutional kernels were 3D and had a shape of 3x3x3. To maximize the flow of information and avoid the vanishing gradient problem [24], previously extracted feature maps are concatenated together with the current output along the feature axis. The convolutional stages are followed by two densely connected layers of sizes 2048 and 512 nodes respectively to try to learn non-linear combinations of the previously extracted features, and a final softmax layer to encode the probability of the two classes of outputs by scaling the activations at the last layer as a summed probability that could be compared to a one-hot encoded vector of the ground truth.

The network was implemented using the the Tensorflow library (version 1.4.0, Google) and was trained and tested on a TitanXp GPU.

### 2.6. Network training

Neural network training is usually done by passing an input signal (e.g. an image) to the network input layer, which in turn, extracts feature maps by convolving kernels with defined shape (e.g. 3x3). These kernels are parametrized by numbers that are called weights. The network output is the result of all operations that happen along the different network's layers. When the output is compared to the ground truth using a loss function, the error is back-propagated so that the network's parameters are updated. AdamOptimizer is used in our

approach to minimize the loss between the predicted output and the ground truth. A learning rate of 10e-5, a mini-batch of size 50 images, and a dropout of rate 0.5 on the densely connected layers were used. All weights were randomly initialized, and L2 norm regularization of $\beta$-value of 10-4 on neuron weights was applied.
The whole dataset was used during training and testing phases in an eightfold cross-validation so that all possible combinations of dataset images were studied. The training time of the network was around 3.2h while the computation time for a single image was 0.26s during the inference phase.
Network performance was tested by plotting the receiver operating characteristic (ROC) curve and calculating the corresponding AUC and sensitivity and specificity.

### 2.7. Overfitting

Overfitting occurs when the model becomes good at classifying the data that is part of the training set yet not as good on the unseen data.

Deeper layers can learn higher level of features and therefore, deeper networks are much easier to train and they are more likely to produce better output performance and higher accuracy [24] [35]. On the other hand, deep networks with the use of a large number of parameters are subject to the problem of information loss and vanishing gradients. Moreover, they are more prone to overfitting, which in turn causes the network to fail in generalizing to an unseen data [36]. To avoid overfitting, we can use one or all of the following methods: first, we can add more data which helps the network to learn more variations of the input and thus generalize better to unseen data. However, this not possible in many applications due to the limitations and difficulty in obtaining more data especially in medical applications. Second, we can use dropout, as one type of a regularization, meaning that random choices of activation are set to zero, which forces the network to find new paths instead of following predetermined patterns. Third, there is another option of regularization, where certain suspiciously large weights are penalized by a regularization function such as the L1 or L2 norm.

$$L' = L + \beta \frac{1}{2} \|W\|^2 \qquad (3)$$

where $L'$ is the regularized loss, $W$ is the weight matrix, and $\beta$ is the penalization coefficient. Lastly, we can use image augmentation, which involves the use of different kinds of image transformation (rigid or elastic) to produce more training images that hold the same information as the original ones, yet are slightly modified so that the network can deal with those images as additional input.
In our study, we used various types of image augmentation, see few examples in Figure 6. First, we shifted the cropping centre around the lesion with values between -7 and +7 mm in x and y direction. After this step, 6 types of additional images augmentation were applied such as flipping around x and y axes, rotation of images in both negative and positive direction around x axis, image brightening, darkening, sharpening and adding noise. This augmentation step resulted in a total number of 12,000 images, which were later split into training and test

sets. It is worth noting that during splitting some of the augmentations for the images with non-significant lesions were not used to ensure the class balance in both training and test sets.

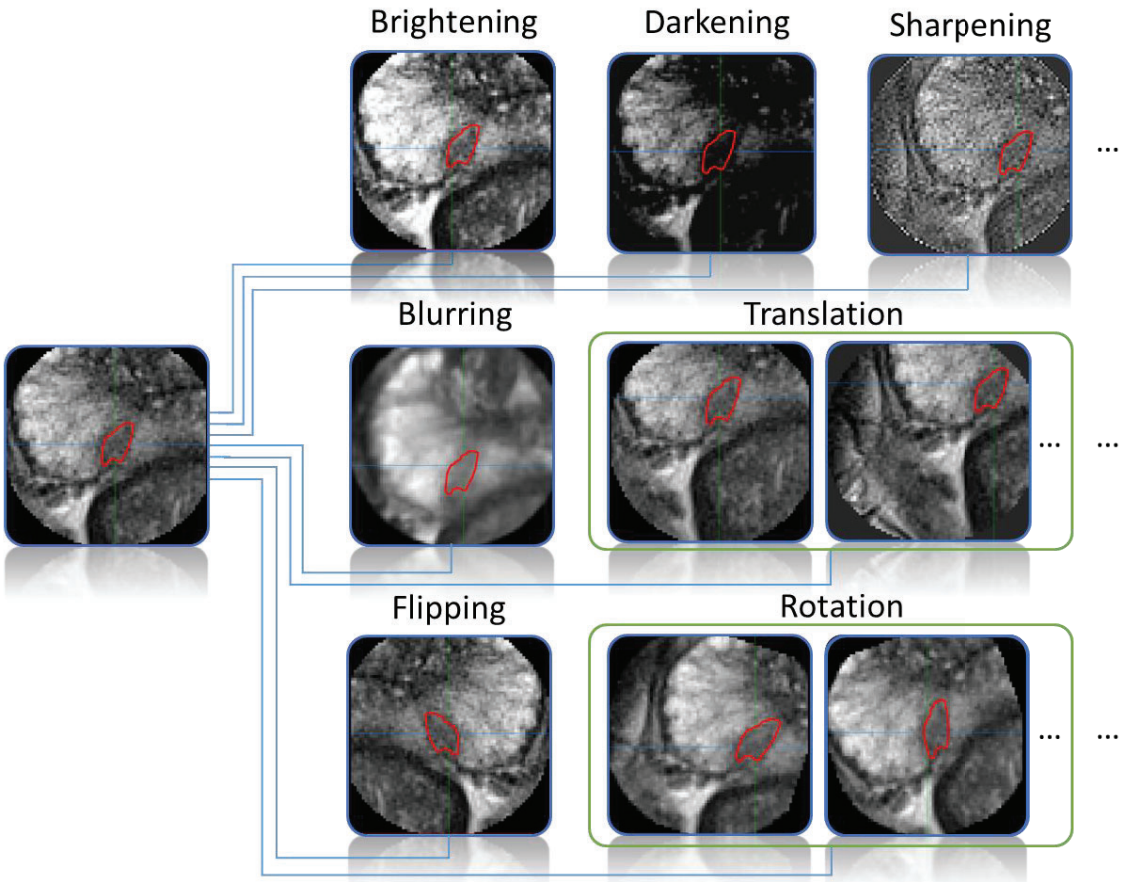


**Figure 6:** Image augmentation: it shows different examples of image augmentation methods that were used to increase the number of training images and avoid overfitting. Red line encircles the suspicious lesion.

## 3. Results

On one eighth of the entire dataset, we tested our developed network while the rest was used for training in an eight-fold cross-validation fashion. Furthermore, we compared our results to other similar studies that were published on the same dataset. Individual pulse sequences or combinations were used as input while biopsy results were used as the ground truth. Compared to the remaining combinations, input combination of ADC, DWI and K-trans (group 2) had the best performance in comparison to the rest of the combinations with 0.897, 81.9% and 86.1% for AUC, sensitivity and specificity respectively, which is comparable to the performance reported for radiologists using PI-RADS V2 [13] and higher than the reported value in [29] and most other similar studies [28] [26] [27] that were based on the same dataset which we used in our study. However, when T2w was added to the previous combination (group 2), the average AUC value for this new combination (group 1) was 0.893, 75.4%, and 92.6% for specificity and sensitivity respectively. The other two combinations (T2w and K-trans (group 3), and T2w, DWI, and ADC (group 4)) had average AUC values of 0.885 and 0.839, respectively. The discrepancies between input groups have been checked for relevance using the t-test. The difference between group 1 and group 2 was not statistically significance with a p-value of 0.25. However, the difference between group 2 and group 3 and between group 2 and group 4 was significant with a p-value of 0.02 and 0.0025, respectively (for details see Figure 7 and table 1).

We chose to compare our results to the ones obtained in the aforementioned studies [29] [28] [26] [27] due to the fact that they used the same dataset which made the comparison more realistic. See Table 4 and discussion section for more details.

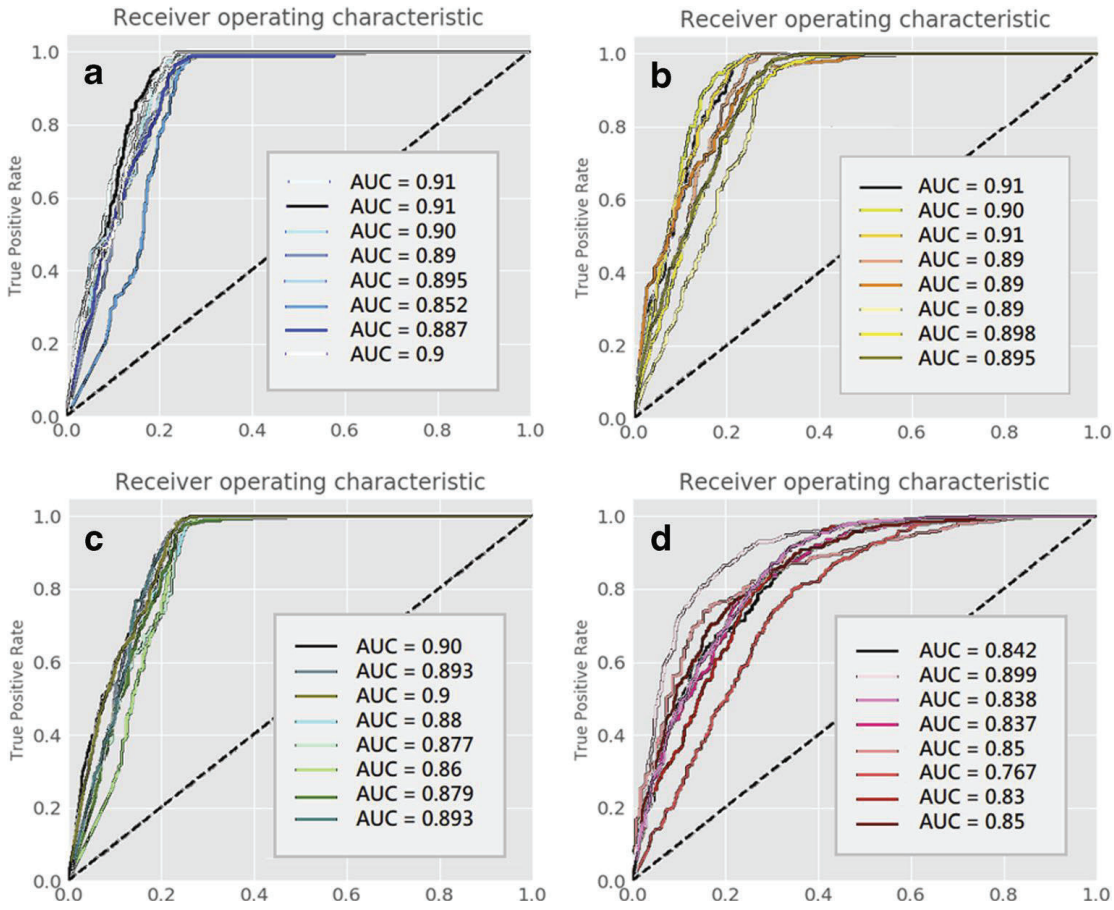


**Figure 7:** ROC curves for assessing network performance for the four main input groups: they show the performance of eight different models resulting from the eightfold cross-validation (colours) of the four combination groups. Each letter denotes the network performance of each combination group: a T2w, ADC, DWI, and K-trans; b ADC, DWI, and K-trans; c T2 and K-trans; d T2, ADC, and DWI. AUC values are provided in the legends. This figure is from Aldoj et al. [34].

Tables 1 presents network performance results for the main four sequence combination groups that were addressed in this study.

**Table 1:** Network performance for each of the eightfold cross-validation on the four main groups of input combinations with 95 % confidence intervals (CI). This Table is from Aldoj et al. [34].

| Group 1: T2w+ADC+DWI+Ktrans | | | | | | |
|---|---|---|---|---|---|---|
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.91 | 0.898 - 0.93 | 82.0% | ± 2.5 % | 90.0 % | ± 2.1% |
| 2 | 0.91 | 0.897 - 0.926 | 76.7 % | ± 3.7 % | 96.0 % | ± 2.0 % |
| 3 | 0.90 | 0.889 - 0.914 | 71.1 % | ± 2.3 % | 97.7 % | ± 2.2 % |
| 4 | 0.89 | 0.873 - 0.909 | 77.2 % | ± 2.8 % | 92.4 % | ± 2.2 % |
| 5 | 0.895 | 0.872 - 0.918 | 87.2 % | ± 2.7 % | 81.6 % | ± 3.3 % |
| 6 | 0.852 | 0.826 - 0.878 | 82.5 % | ± 2.6 % | 89.5 % | ± 2.8 % |
| 7 | 0.887 | 0.873 - 0.903 | 61.5 % | ± 3.2 % | 97.2 % | ± 1.9% |
| 8 | 0.9 | 0.891 - 0.92 | 65.5 % | ± 3.5 % | 97.0 % | ± 1.8 % |
| Average AUC | 0.893 | | 75.4 % | | 92.6 % | |
| Standard deviation | 0.018 | | 0.08 | | 0.05 | |
| Group 2: ADC+DWI+Ktrans | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.91 | 0.890 - 0.923 | 81.2 % | ± 2.9 % | 90.5 % | ± 3.2 % |
| 2 | 0.90 | 0.898 - 0.920 | 75.8 % | ± 2.5% | 9.1 % | ± 1.7 % |
| 3 | 0.91 | 0.899 - 0.923 | 71.2 % | ± 3.2 % | 97.1 % | ± 2.3 % |
| 4 | 0.89 | 0.873 - 0.908 | 77.2 % | ± 3.6 % | 92.00% | ± 2.8 % |
| 5 | 0.89 | 0.874 - 0.916 | 87.60% | ± 2.2 % | 80.00% | ± 2.1 % |
| 6 | 0.89 | 0.874 - 0.916 | 87.4 % | ± 2.5 % | 79.3 % | ± 2.9 % |
| 7 | 0.898 | 0.878 - 0.918 | 87.6 % | ± 2.4 % | 79.7 % | ± 4.2 % |
| 8 | 0.895 | 0.874 - 0.916 | 87.6 % | ± 2.3 % | 79.1 % | ± 2.5 % |
| Average AUC | 0.897 | | 81.9% | | 86.1 % | |
| Standard deviation | 0.008 | | 0.06 | | 0.07 | |
| Group 3: T2w +Ktrans | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.90 | 0.886 - 0.92 | 80.5 % | ± 2.5 % | 89.5 % | ± 2.2 % |
| 2 | 0.893 | 0.877 - 0.91 | 74.4 % | ± 3.3% | 89.3 % | ± 2.5 % |
| 3 | 0.90 | 0.886 - 0.92 | 80.4 % | ± 3.3 % | 89.9 % | ± 1.2 % |
| 4 | 0.88 | 0.866 - 0.902 | 76.0 % | ± 3.7 % | 95.0 % | ± 2.3 % |
| 5 | 0.877 | 0.851 - 0.902 | 87.7 % | ± 2.2 % | 84.4 % | ± 2.5 % |
| 6 | 0.86 | 0.833 - 0.885 | 82.9 % | ± 3.0 % | 93.7 % | ± 2.4 % |
| 7 | 0.879 | 0.865 - 0.895 | 60.7 % | ± 3.6 % | 95.9 % | ± 2.2 % |
| 8 | 0.893 | 0.877 - 0.909 | 63.0 % | ± 2.9% | 97.5 % | ± 1.7 % |
| Average AUC | 0.885 | | 75.7% | | 91.9 % | |
| Standard deviation | 0.013 | | 0.09 | | 0.04 | |
| Group 4: T2w+ADC+DWI | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.842 | 0.823 - 0.863 | 72.7 % | ± 2.8 % | 79.5 % | ± 2.9 % |
| 2 | 0.899 | 0.884 - 0.916 | 71.9 % | ± 2.8 % | 91.5 % | ± 2.5 % |
| 3 | 0.838 | 0.818 - 0.859 | 73.0 % | ± 3.2 % | 80.4 % | ± 2.5% |
| 4 | 0.837 | 0.817 - 0.859 | 72.9 % | ± 3.0 % | 80.8 % | ± 2.2 % |
| 5 | 0.85 | 0.827 - 0.872 | 86.0 % | ± 3.1 % | 66.7 % | ± 3.9 % |
| 6 | 0.767 | 0.739 - 0.797 | 73.7 % | ± 3.5 % | 72.1 % | ± 3.5 % |
| 7 | 0.83 | 0.815 - 0.852 | 52.1 % | ± 4.3 % | 92.5 % | ± 2.5 % |
| 8 | 0.85 | 0.832 - 0.872 | 55.8 % | ± 3.9 % | 90.2 % | ± 2.3 % |
| Average AUC | 0.839 | | 69.7% | | 81.7% | |
| Standard deviation | 0.036 | | 0.10 | | 0.09 | |

Tables 2 and 3 present network performance results for the use of pairs of MR sequence combinations and for the use of single MR sequence.

**Table 2:** Network performance for all eightfold cross-validations on pairwise combinations of MR sequences as input with 95 % confidence intervals (CI). This Table is from Aldoj et al. [34].

| Group 5: DWI+Ktrans | | | | | | |
|---|---|---|---|---|---|---|
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.90 | 0.885 - 0.915 | 81.8 % | ± 2.8 % | 90.2% | ± 3 % |
| 2 | 0.90 | 0.888 - 0.915 | 84 .3% | ± 2.9 % | 97.0 % | ± 2 % |
| 3 | 0.87 | 0.866 - 0.891 | 67.6 % | ± 2.9 % | 96.7 % | ± 1.9 % |
| 4 | 0.86 | 0.848 - 0.883 | 77.0 % | ± 3 % | 94.2 % | ± 2.2 % |
| 5 | 0.86 | 0.84 - 0.883 | 87.8 % | ± 2.5 % | 75.7 % | ± 5 % |
| 6 | 0.84 | 0.818 - 0.864 | 84.5 % | ± 3 % | 86.8 % | ± 4.2 % |
| 7 | 0.867 | 0.853 - 0.882 | 61.1 % | ± 3.5 % | 93.0 % | ± 1.8% |
| 8 | 0.877 | 0.863 - 0.892 | 63.8 % | ± 3.6 % | 96.6 % | ± 1.5 % |
| Average AUC | 0.87 | | 76 % | | 91.2 % | |
| Standard deviation | 0.02 | | 0.10 | | 0.07 | |
| Group 6: ADC+Ktrans | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.90 | 0.885 - 0.915 | 85.0 % | ± 3 % | 82.0 % | ± 2% |
| 2 | 0.88 | 0.861 - 0.889 | 74 % | ± 2.5 % | 93.4 % | ± 1.8 % |
| 3 | 0.87 | 0.866 - 0.891 | 67.1 % | ± 3.8 % | 93.8 % | ± 2.4 % |
| 4 | 0.853 | 0.841 - 0.863 | 75.6 % | ± 3 % | 90.0 % | ± 2.2 % |
| 5 | 0.85 | 0.838 - 0.878 | 87.5 % | ± 2.2 % | 75.3 % | ± 2 % |
| 6 | 0.83 | 0.82 - 0.852 | 56.3 % | ± 3.8 % | 90.7 % | ± 2.3 % |
| 7 | 0.833 | 0.812 - 0.854 | 56.8 % | ± 3.4 % | 91.2 % | ± 2.4 % |
| 8 | 0.87 | 0.860 - 0.891 | 59.8 % | ± 3.9 % | 95.4 % | ± 2 % |
| Average AUC | 0.86 | | 70.2 % | | 88.9 % | |
| Standard deviation | 0.02 | | 0.12 | | 0.07 | |
| Group 7: ADC+DWI | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.816 | 0.795 - 0.839 | 78.9 % | ± 2.7 % | 73.7 % | ± 3.5 % |
| 2 | 0.822 | 0.81 - 0.833 | 74.2 % | ± 3.1% | 88.6 % | ± 2.3 % |
| 3 | 0.80 | 0.789 - 0.813 | 59.9 % | ± 3 % | 85.6 % | ± 1.3 % |
| 4 | 0.77 | 0.758 - 0.79 | 67.7 % | ± 4.1 % | 74.3 % | ± 2.6 % |
| 5 | 0.754 | 0.74 - 0.769 | 82.9 % | ± 2.6 % | 52.5 % | ± 3.9 % |
| 6 | 0.83 | 0.806 - 0.855 | 83.4 % | ± 3.1 % | 79.7 % | ± 3.5 % |
| 7 | 0.73 | 0.718 - 0.74 | 77.1 % | ± 3.5 % | 57.0 % | ± 3.2 % |
| 8 | 0.75 | 0.739 - 0.762 | 81.8 % | ± 2.7 % | 52.0 % | ± 4.4 % |
| Average AUC | 0.784 | | 74.8% | | 70.4 % | |
| Standard deviation | 0.03 | | 0.07 | | 0.14 | |

**Table 3:** Network performance results for all eightfold cross-validation on individual MR sequences s input with 95 % confidence intervals (CI). This Table is from Aldoj et al. [34].

| Group 8: T2 | | | | | | |
|---|---|---|---|---|---|---|
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.70 | 0.684 - 0.731 | 67.7 % | ± 3.2 % | 60.8 % | ± 3.7 % |
| 2 | 0.82 | 0.811 - 0.847 | 72.7 % | ± 2.8 % | 80.0 % | ± 3.7 % |
| 3 | 0.72 | 0.704 - 0.743 | 55.7 % | ± 3.4 % | 77.3 % | ± 2.9 % |
| 4 | 0.726 | 0.701 - 0.749 | 66.7 % | ± 3.2 % | 68.8 % | ± 2.6 % |
| 5 | 0.716 | 0.688 - 0.744 | 81.3 % | ± 3.4 % | 47.1 % | ± 4.4 % |
| 6 | 0.71 | 0.684 - 0.736 | 74.7 % | ± 3.4 % | 56.6 % | ± 3.8 % |
| 7 | 0.715 | 0.690 - 0.736 | 45.9 % | ± 4.2 % | 80.1 % | ± 2.6 % |
| 8 | 0.77 | 0.750 - 0.791 | 52.0 % | ± 3.3 % | 82.9 % | ± 2.1 % |
| Average AUC | 0.734 | | 64.58% | | 69.2% | |
| Standard deviation | 0.04 | | 0.12 | | 0.13 | |
| Group 9: ADC | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.77 | 0.757 - 0.8 | 76.5 % | ± 3 % | 65.2 % | ± 3.2 % |
| 2 | 0.786 | 0.773 - 0.812 | 69.6 % | ± 2.5 % | 74.4 % | ± 1.8 % |
| 3 | 0.727 | 0.706 - 0.744 | 53.1 % | ± 3.8 % | 77.8 % | ± 3.1 % |
| 4 | 0.703 | 0.679 - 0.728 | 62.4 % | ± 3 % | 67.3 % | ± 3.2 % |
| 5 | 0.75 | 0.722 - 0.774 | 83.0 % | ± 2.2 % | 53.4 % | ± 3.4 % |
| 6 | 0.70 | 0.675 - 0.728 | 75.4 % | ± 3.2 % | 54.8 % | ± 4.0 % |
| 7 | 0.683 | 0.661 - 0.708 | 44.5 % | ± 4.4 % | 77.6 % | ± 2.9 % |
| 8 | 0.77 | 0.749 - 0.793 | 53.4 % | ± 3.9 % | 84.6 % | ± 2 % |
| Average AUC | 0.736 | | 64.7 % | | 69.4 % | |
| Standard deviation | 0.03 | | 0.13 | | 0.11 | |
| Group 10: DWI | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.808 | 0.789 - 0.828 | 76.5 % | ± 2.7 % | 72.2 % | ± 3.2 % |
| 2 | 0.817 | 0.800 - 0.836 | 71.1 % | ± 3.1% | 81.8 % | ± 2.3 % |
| 3 | 0.815 | 0.799 - 0.831 | 60.4 % | ± 3 % | 87.3 % | ± 1.3 % |
| 4 | 0.80 | 0.787 - 0.827 | 72.3 % | ± 3.1 % | 76.0 % | ± 2.6 % |
| 5 | 0.795 | 0.769 - 0.818 | 87.8 % | ± 2.6 % | 56.4 % | ± 4.3 % |
| 6 | 0.742 | 0.717 - 0.767 | 76.9 % | ± 3.1 % | 60.7 % | ± 3.5 % |
| 7 | 0.807 | 0.789 - 0.825 | 57.5 % | ± 3.5 % | 85.9 % | ± 2 % |
| 8 | 0.81 | 0.797 - 0.830 | 60.1 % | ± 2.7 % | 87.0 % | ± 1.4 % |
| Average AUC | 0.79 | | 70.3 % | | 75.9 % | |
| Standard deviation | 0.02 | | 0.10 | | 0.12 | |
| Group 11: Ktrans | | | | | | |
| Model number | AUC | CI (95) | Sensitivity | ± CI 95% | Specificity | ± CI 95% |
| 1 | 0.82 | 0.799 - 0.838 | 76.6 % | ± 2.7 % | 74.4 % | ± 3.1 % |
| 2 | 0.80 | 0.782 - 0.821 | 63.8 % | ± 3.1% | 82.7 % | ± 2.2 % |
| 3 | 0.80 | 0.780 - 0.814 | 51.8 % | ± 4 % | 87.6 % | ± 1.5 % |
| 4 | 0.80 | 0.778 - 0.818 | 70.3 % | ± 3.1 % | 79.6 % | ± 2.6 % |
| 5 | 0.81 | 0.779 - 0.83 | 84.1 % | ± 2.1 % | 66.4 % | ± 3.3 % |
| 6 | 0.793 | 0.768 - 0.82 | 80.2 % | ± 3.1 % | 73.1 % | ± 3.5 % |
| 7 | 0.79 | 0.765 - 0.808 | 49.5 % | ± 4.1 % | 87.7 % | ± 1.4 % |
| 8 | 0.77 | 0.75 - 0.79 | 52.1 % | ± 3.7 % | 72.0 % | ± 2.4 % |
| Average AUC | 0.798 | | 66.0 % | | 77.9 % | |
| Standard deviation | 0.02 | | 0.13 | | 0.07 | |

Tables 4 presents the comparison of our network's performance against other studies that were published using the same dataset, and highlights the similarities and the differences in data handling, input dimensions, size of training set and the type of used algorithm.

**Table 4:** Comparison between our approach and previously reported studies.

| Author | Study Objective | Dataset Size | Architecture | Input's dimension | Preprocessing | AUC | Data Source |
|---|---|---|---|---|---|---|---|
| **Kiraly et al [28]** | Detection/ classification | 202 | Modified SegNet | 2D<br>3D | 3D elastic registration<br>3D Gaussian kernel | 0.834 | ProstateX |
| **Chen et al [26]** | Classification | 344 | InceptionV3<br>VGG-16 | 2D | Data augmentation<br>Normalization<br>Compressing images into a three-channel RGB | 0.81<br>0.83 | ProstateX |
| **Liu et al [37]** | Classification | 336 | XmasNet | 3D | Linear interpolation<br>Co-registration<br>Refining lesion center<br>Creating four inputs as RGB channels<br>Data augmentation | 0.84 | ProstateX |
| **Mehrtash et al [29]** | Classification | 341 | Three parallel pipeline Network | 3D | Image resampling<br>Image cropping<br>Data augmentation<br>Normalization | 0.80 | ProstateX |
| **Aldoj et al [34]** | Classification | 200 | Single pipeline network | 3D | Image resampling<br>Spherical cropping<br>Data augmentation<br>Normalization | 0. 897 | ProstateX |
| **Song et al [27]** | Classification | 195 | Modified VGG-Net | 2D | Registration<br>ROI labeling<br>Data augmentation<br>Patch extraction<br>Normalization | 0.944 | ProstateX |

The best model from group 2 has been used in any study and test discussed below, and unless otherwise specified, the images and results described are from this model. This model had an AUC of 0.91 with 81.2% sensitivity and 90.5% specificity. Figure 8 presents four different test cases as examples of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions of the network.

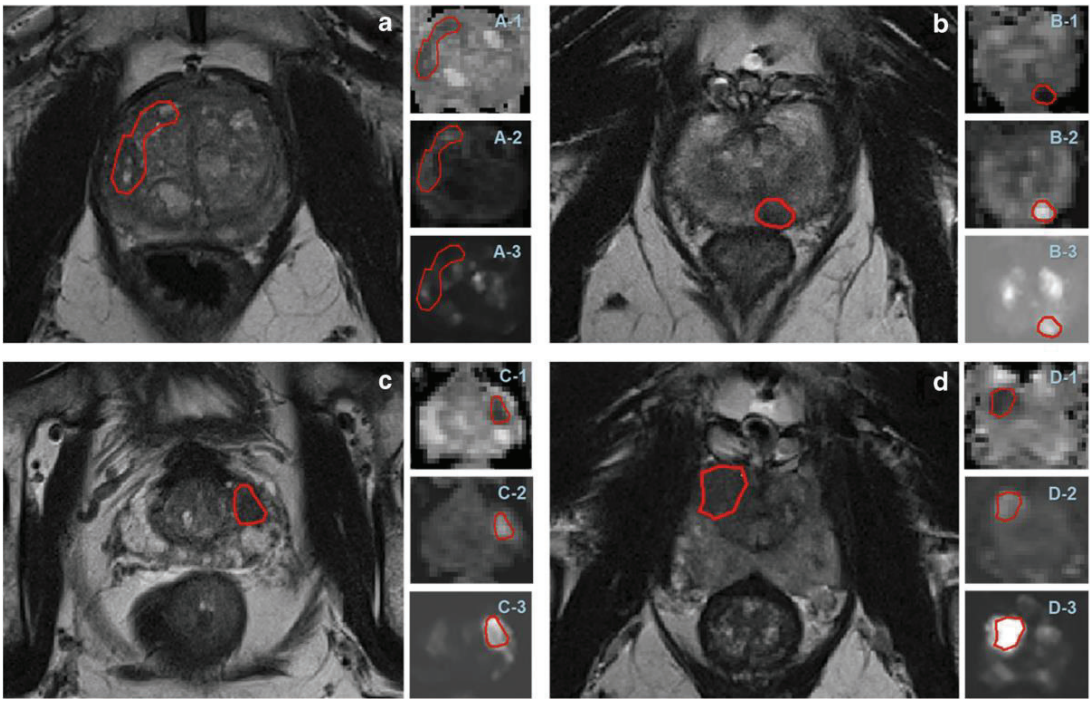


**Figure 8:** Image examples to illustrate network performance: a. true positive; b. false positive; c. false negative, and d. true negative. The suspicious lesion is encircled by red line. This figure is from Aldoj et al. [34].

The comparative overview of lesion volumes and largest lesion diameters is shown in Figure 9 as box plots of the mean and median volumes and diameters with their respective statistical groups. This figure helps to investigate whether or not the size of the lesion has an impact on the performance of the network. It is obviously shown that the lesion size has neither a negative nor a positive impact on the network performance.

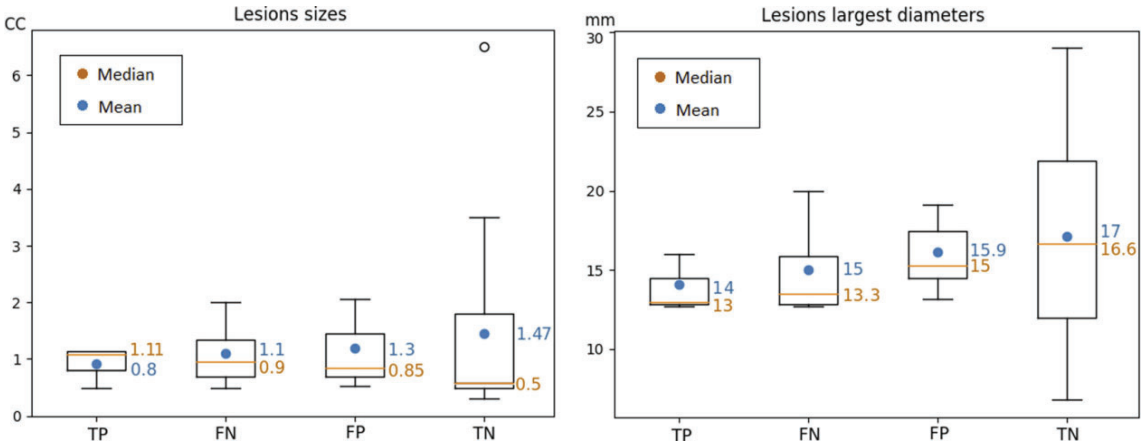


**Figure 9:** Boxplot of the sizes and largest diameter of the lesions. It shows lesion sizes on the left figure and largest diameter of the right for all four statistical categories. Numbers with orange represent the mean value while the number with blue represent the median. This figure is from Aldoj et al. [34].

Figure 10 demonstrates and helps to track the network performance during the training and test process, and illustrates the training and test accuracy performance throughout the training period. The network was able to learn and increase accuracy over time, as seen in the

figure on the left side. The network reached a point about the 4000 iteration step where the accuracy no longer increased, so the training was halted. The figure on the right side shows the individual accuracy with respect to lesions locations, which reflects, in general, similar distribution as the original test set that was shown in figure 3.

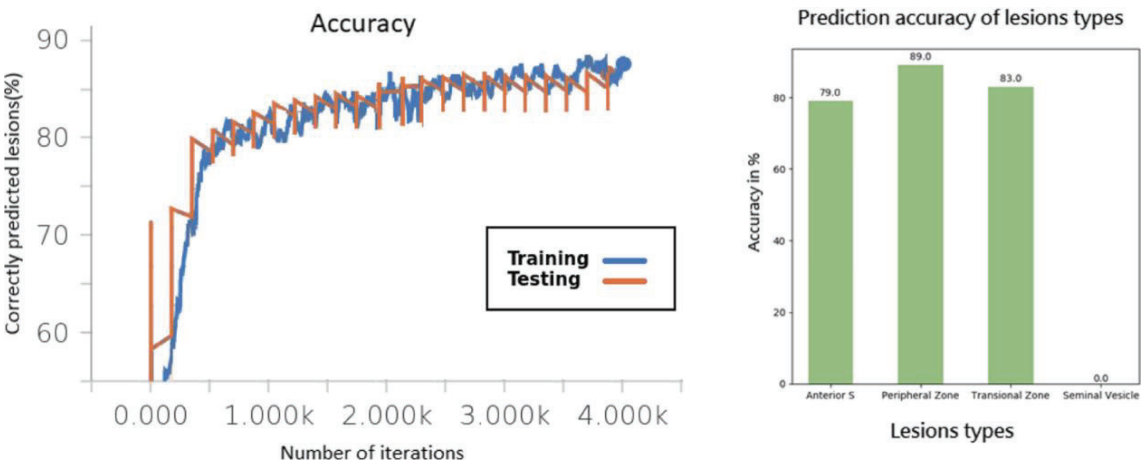


**Figure 10**: The curves on the left represent the accuracy of the network during training and testing along with iteration time. The bar graph on the right displays network accuracy for different lesion locations. This figure is from Aldoj et al. [34].

Finally, examples of lesions wrongly classified by the network for all types of examined MRI sequences are shown in Figure 11. In all the MR sequences used it demonstrates multiple forms of lesions with the predictions of their corresponding network. This figure helps to investigate why the network predicted some cases incorrectly.

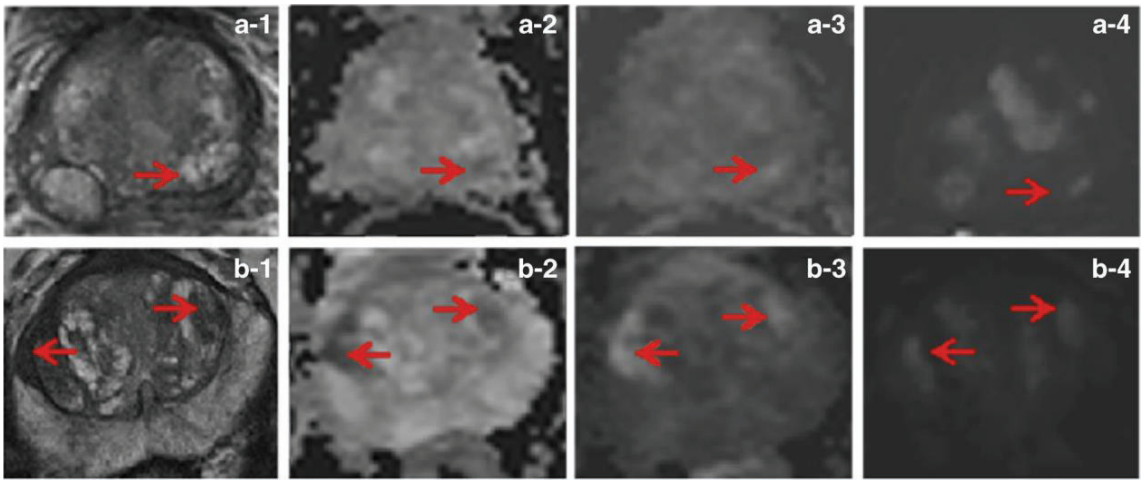


**Figure 11:** Illustration of incorrectly predicted lesions: a. Network classifies a nonsignificant lesion as significant. b. Patient with two lesions – a significant lesion on the right and a nonsignificant lesion on the left – and the network misclassifies both lesions Arrows in the figure indicate the rough position of the lesion, and numbers stand for: 1 = T2w, 2 = ADC, 3 = DWI, and 4 = K-trans. This figure is from Aldoj et al. [34].

# 4. Discussion

Men who are suspected to have a significantly prostatic lesions are scanned with mp-MRI. Radiologists study their radiographs and decide for further clinical actions. Deep learning can have a great potential in assisting the diagnostic and evaluation procedures that radiologists usually handle manually.

In this work, we developed a 3D CNN-based approach to predict the probability of a given prostate lesion to be clinically significant using mp-MRI datasets. As an input, we used a 3D multimodal cropped regions around the suspected lesion and passed them as an input into our CNN where a final classification layer was represented by two output neurons of a Softmax function. Our network was tested using the images included in this dataset in different combinations or alone to assess how this affected network performance in classifying prostate lesions into significant and nonsignificant cancers. Biopsy-based histological diagnosis acted as our ground truth during our network's training and testing.

In order to investigate the impact of each MR sequence on the output of the network, specifically, multiple input combinations were evaluated and analyzed using the same model architecture. In terms of average AUC, sensitivity and accuracy, the input used in group, which includes ADC, DWI and K-trans images, resulted in the best network output. In addition, the findings given in Table 1 clearly demonstrate that adding T2w images to the network input has the least effect on enhancing the performance of the network. This inference is based on a p-value of 0.25 that is greater than 0.05 between groups 1 and 2, and thus the difference is not important. The fact that the T2w sequence contains only anatomical details that could suggest the location of a suspicious tissue, yet no functional information about the nature and characteristics of the tissue, may support this finding.

The heterogeneity in the tissue which led to suspected appearance could be just a normal variety in the tissue structure or BPH and does not indicate any cancerous characteristics. In contrast to group 1, the p-value is significant at 0.02 when comparing groups 2 and 3, which suggests that the influence of DWI (b-value and ADC) is relevant to network performance. Similarly, by comparing groups 2 and 4, the inclusion of K-trans results in significant performance enhancement and results in a p-value of 0.0025. This suggests that k-trans has the greatest influence on the efficiency of the network and turns out to be the MR sequence that better represents the characteristics of prostate lesions. These conclusions are supported by the results of more detailed investigation of other combinations of inputs (presented in Table 2) or standalone sequences (Table 3) where it shows that T2 was the worst performing input among the other sequences while the rest of the sequences, since they carry functional information, were better in terms of tumour characterisation and hence better diagnostic performance. K-trans is the highest in terms of its importance for lesion characterization (significant or non-significant) and is preceded by DWI and ADC. T2w, on the other hand, has the least diagnostic impact and is thus the least significant when it comes to classifying tumors.

### 4.1. Comparison with previous studies

Earlier studies focused mostly on tackling the classification problem using 2D networks with either a proper slice selection or segmented region of the suspected lesion as input [22] [25]. When it comes to PROSTATEx dataset, few studies tested their 2D/3D network on it (comparison can be seen in Table 4), where our approach was based on 3D image without any segmentation or slice selection. Kiraly et al. [28] proposed an approach that investigated both detecting the lesions and classifying their aggressiveness by using a convolutional encoder-decoder. The task of detection was applied through semantic segmentation. Their obtained average AUC was 0.834 for the classification. On the other hand, Chen et al [26] investigated the possibility of using transfer learning of InceptionV3 and VGG-16 model that were originally trained on ImageNet dataset, while Song et al [27] used a patch-based approach based on VGGNet with 131 layers. The achieved AUCs were 0.81, 0.83 and 0.944 for inceptionV3, and VGGNet from Chen et al and Song et al respectively. However, In contrast to our approach the aforementioned methods were 2D and required lengthy image preparation and careful slice selection, which precludes their use in daily practice. Thus, designing a 3D tool that requires minimal manual preparation is highly desirable. Liu et al [37] developed a CNN termed XmasNet which was trained and tested on PROSTATEx challenge and achieved an AUC of 0.84, yet it also required a pre-segmentation of lesion as the network only trained on suspected tissues. Another study by Mehrtash et al [29] addressed the classification problem in a similar way as we did. They designed a network with 3 parallel pipelines and achieved an AUC of 0.80. In contrast to their method, our approach consisted of a single pipeline (which made it more compact, parameters efficient and easier to train) that processed all input sequences at the same time and required the least amount of preparation, which involved only image resampling and lesion localization. Furthermore, Our approach achieved a performance in terms of AUC, sensitivity and specificity very similar to the results reported in to [29] [37] and to the performance reported for an experienced reader [13].

### 4.2. Clinical implications

In clinical practice, a model should have high values of AUC, sensitivity, and specificity, yet, sensitivity is more important than specificity since the more important aim is to lower the false negative rate, where the significant lesion is miss-diagnosed as non-significant and hence, the tumour is missed and left untreated which causes the tumour to keep growing and spreading. It might be not a realistic vision in the near future to have a reliable and robust approach that can fully replace the radiologist in specific or general diagnostic tasks. However, the designed model (s) can augment and assist the radiologists in their daily clinical practice and carries some parts of the heavy loads when dealing with large amount of imaging volumes.

An experienced reader using PI-RADS v2 in interpreting mp-MRI datasets is reported to have an AUC of 0.83, sensitivity of 77%, and specificity of 81% [13]. While Lui et al achieved an average AUC of 0.84 and Mehrtash et al an AUC of 0.8, our model had a higher AUC (model 1 from group 2) than these studies and [37], yet it was less sensitive and more specific than [37]. While the outcomes of this model seem to be promising, it is not advanced enough to substitute the radiologists by any means. By corroborating their decision, though, radiologists

will benefit from the superiority of the model in terms of specificity and increase its specificity for improved diagnostic results in patients.

### 4.3. Explainability

In relation to the efficacy of an AI-based diagnostic tool, one of the main things discussed is explainability, which implies the ability of the tool to interpret how and on the basis of what a certain diagnostic judgment was taken, and why this diagnosis in the classification case was accomplished. Deep convolutional neural networks have shown an impressive performance in many image classification tasks. However, because of their nature of multi-layer architecture, they are considered as black box and it is hard to explain how they arrive at a specific prediction. Many studies showed some examples of correct object classification outputs that were caused by some activations that were resulted from wrong part of the input image such as image captions [38].

Explainability may be superimposed on the original picture in the form of heat maps, which in turn show the location of the suspicious lesion and how confident the model is in predicting outcomes based on the examination of lesions. In other words, the heat map highlights the impact of individual pixel (voxel) in the image with respect to the prediction output of the network and help to understand how the network arrive at that particular outcome. In case of 2D models, such explainability maps can be provided by various methods and algorithms (e.g. layer-wise relevance propagation [38], activation maps [39] etc.) that provide information on which part of the image contributed positively or negatively to the outcome and by how much. All these and similar methods are optimized to be used on 2D images. In the case of 3D, however, there is currently no accurate algorithm that can extract and visualize similar information in 3D heat maps. In addition, the development of a 3D algorithm to create maps of explainability is not an easy problem to solve and is actually beyond the scope of this study and should be addressed in a separate project.

### 4.4. CNN performance and design considerations

Network performance is usually affected by many factors such as intrinsic network features (number of layers, parameters etc.) and the quality of input images. However, it is highly unlikely that a network surpasses the values obtained with the experienced reader according to PI-RADS v2 in similar realistic clinical setting where the images are presented in their plain 3D format without prior segmentation or lesion detection [13]. To monitor the network's performance, several values such as AUC, sensitivity and specificity were reported, see Figure 7. Additionally, we plotted the accuracy curve, see Figure 10, of correctly predicted cases over time to show how the network is performing across the training and test phases.

As mentioned earlier, several aspects and hyperparameters can influence the predicted results of a certain network, such as depth of the network, skip connections, choice of the loss function, learning rate value, the choice of regularization method etc.

The deeper the network, the larger the receptive field and hence, the greater its capability to extract more information from the images. However, deeper networks suffer the problem of

large number of parameters where the problem of overfitting becomes more apparent. Therefore, the balance between the model with sufficient abstract information and the model with suitable depth and number of parameters should be considered [27].

Pre-trained networks proved to be performing better and achieving higher accuracy values than the networks trained from scratch [22]. However, all available pre-trained networks had 2D-based architectures and could not be used in our study since the input images were 3D and therefore, training from scratch was the only choice.

Figure 11 presents two examples from our dataset that were incorrectly classified by our network, each row shows a different case and each column represents a different MR sequence. Case A shows a patient with a nonsignificant lesion, yet it was classified incorrectly as a significant lesion. This could be due to the fact that this lesion was present in all sequences and had a higher contrast in comparison to its surrounding tissue, which in turn highlighted the lesion and strengthened the activation signal inside the network toward false classification and hence a false positive prediction. In patient B, two lesions were present, a significant one in the left and a nonsignificant lesion in the right prostate. The network classified both lesion incorrectly. This is attributed to the relatively same level of contrast between the lesion and the adjacent tissues and its presence in all MR sequences. Furthermore, our network is not perfect and it has a margin of errors, which needs to be addressed in future work.

### 4.5. Effect of lesion size

When looking at Figure 8, one might think that classification of prostate lesions is influenced by its size or diameter and that larger lesions might help the network to achieve better performance and accuracy since they are harder to miss and the their pixels (voxels) signal is much stronger because it is resulted from a larger part of the image. However, Figure 9 illustrates that in our approach and used dataset there is no relationship between lesion size or diameter, and network performance. Hence, there is no positive or negative correlation between lesion size and network accuracy. Therefore, larger or smaller lesions sizes have no significant effect on the overall results.

## 5. Study limitations

As a training and test dataset, only 200 patients (318 lesion volumes) were used, which is considered small for training a neural network, which is considered a data-hungry method and thousands or hundreds of thousands of images are required to be used in many implementations. Although the 318 volumes were augmented many times using different types of methods of image transformation to expand the training examples and address overfitting, this amount is still not adequate and further images are required to obtain improved output. Let alone that the resulted augmented images are just different variations of the original ones and they do not replace the importance of new images which contain important clinical data.

Furthermore, we did not use complete images. Rather, our network classified prostate lesions as significant or not based on a small input volume. This, in turn, does not reflect the realistic

clinical routine where images are handled and investigated by the radiologist as they are resulted from the scanner without prior segmentation, delineation or slice selection. Finally, the network only made diagnostic predictions but yielded no localization or explainability maps, which is highly desirable for robustness and reliability. Therefore further research is required to address current constraints and enhance the efficiency of our network along these lines and allow its clinical routine use.

## 6. Future work

Several aspects are needed to be addressed in future work to increase the applicability of such approaches. For instance, a bigger (and preferably multi-centre) dataset is of a high importance and would add important values to our existing work toward robustness and generalizability. Explainability maps are much needed in any diagnostic AI-based tool, thus providing such maps is essential in such approaches. Additionally, different state-of-the-art classification networks should be extended to adapt 3D input, tested in such diagnostic problems and compared to the developed approach. Furthermore, an ensemble of different network architectures could also be considered so that the overall performance is more robust and reliable.

## 7. Conclusion

The preliminary results we achieved with our algorithm suggest that automated classification of prostate cancer using the deep convolutional neural network developed here is feasible and promising. Our developed model requires the least amount of manual work where only the lesion location is required. Our network takes a combination or individual MR sequences (T2w, ADC, DWI, and K-trans) as input and classifies the lesion as significant or nonsignificant. The diagnostic performance of our network, which is quantified by AUC, sensitivity and specificity, is comparable to that reported for human readers and other similar published studies, which makes this tool very promising and demonstrates its great potential for improving the capabilities of deep learning-based diagnosis in prostate cancer classification and making it suitable for routine clinical use.

# 8. References


1 Barentsz , J., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., and Fütterer, J.: 'ESUR prostate MR guidelines', Eur Radiol, 2012, pp. 2012;2022(2014):2746–2757.

2 Vargas, H., Hötker, A., Goldman, D., Moskowitz, C., Gondo, T., Matsumoto, K., Ehdaie, B., Woo, S., Fine, S., and Reuter, V.: 'Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference', Eur Radiol, 2016, 26, (6), pp. 1606-1612

3 Siegel, R.L., Miller, K.D., and Jemal, A.: 'Cancer statistics, 2016', CA: a cancer journal for clinicians, 2016, 66, (1), pp. 7-30

4 Panebianco, V., Barchetti, F., Barentsz, J., Ciardi, A., Cornud, F., Futterer, J., and Villeirs, G.: 'Pitfalls in interpreting mp-MRI of the prostate: a pictorial review with pathologic correlation', Insights into imaging, 2015, 6, (6), pp. 611-630

5 Thompson, I.M., Pauler, D.K., Goodman, P.J., Tangen, C.M., Lucia, M.S., Parnes, H.L., Minasian, L.M., Ford, L.G., Lippman, S.M., and Crawford, E.D.: 'Prevalence of prostate cancer among men with a prostate-specific antigen level≤ 4.0 ng per milliliter', New England Journal of Medicine, 2004, 350, (22), pp. 2239-2246

6 Schröder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., and Zappa, M.: 'Screening and prostate-cancer mortality in a randomized European study', New England Journal of Medicine, 2009, 360, (13), pp. 1320-1328

7 Descotes, J.-L.: 'Diagnosis of prostate cancer', Asian journal of urology, 2019, 6, (2), pp. 129-136

8 Epstein, J.I., Allsbrook Jr, W.C., Amin, M.B., Egevad, L.L., and Committee, I.G.: 'The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma', The American journal of surgical pathology, 2005, 29, (9), pp. 1228-1242

9 Valerio, M., Donaldson, I., Emberton, M., Ehdaie, B., Hadaschik, B.A., Marks, L.S., Mozer, P., Rastinehad, A.R., and Ahmed, H.U.: 'Detection of clinically significant prostate cancer using magnetic resonance imaging–ultrasound fusion targeted biopsy: a systematic review', European urology, 2015, 68, (1), pp. 8-19

10 Villers, A., Lemaitre, L., Haffner, J., and Puech, P.: 'Current status of MRI for the diagnosis, staging and prognosis of prostate cancer: implications for focal therapy and active surveillance', Current opinion in urology, 2009, 19, (3), pp. 274-282

11 Fehr, D., Veeraraghavan, H., Wibmer, A., Gondo, T., Matsumoto, K., Vargas, H.A., Sala, E., Hricak, H., and Deasy, J.O.: 'Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images', Proceedings of the National Academy of Sciences, 2015, 112, (46), pp. E6265-E6273

12 Peng, Y., Jiang, Y., Yang, C., Brown, J.B., Antic, T., Sethi, I., Schmid-Tannwald, C., Giger, M.L., Eggener, S.E., and Oto, A.: 'Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study', Radiology, 2013, 267, (3), pp. 787-796

13 Kasel-Seibert, M., Lehmann, T., Aschenbach, R., Guettler, F.V., Abubrig, M., Grimm, M.-O., Teichgraeber, U., and Franiel, T.: 'Assessment of PI-RADS v2 for the detection of prostate cancer', European journal of radiology, 2016, 85, (4), pp. 726-731

14 Sidhu, H.S., Benigno, S., Ganeshan, B., Dikaios, N., Johnston, E.W., Allen, C., Kirkham, A., Groves, A.M., Ahmed, H.U., and Emberton, M.: 'Textural analysis of multiparametric MRI detects transition zone prostate cancer', Eur Radiol, 2017, 27, (6), pp. 2348-2358

15 Tiwari, P., Kurhanewicz, J., and Madabhushi, A.: 'Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS', Medical image analysis, 2013, 17, (2), pp. 219-235

16 LeCun, Y., Bengio, Y., and Hinton, G.: 'Deep learning', nature, 2015, 521, (7553), pp. 436-444

17 Simonyan, K., and Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:1409.1556, 2014
18 Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q.: 'Densely connected convolutional networks', (2017, edn.), pp. 4700-4708
19 Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R.M.: 'Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning', IEEE transactions on medical imaging, 2016, 35, (5), pp. 1285-1298
20 Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., and Liang, J.: 'Convolutional neural networks for medical image analysis: Full training or fine tuning?', IEEE transactions on medical imaging, 2016, 35, (5), pp. 1299-1312
21 Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., and Rueckert, D.: 'Unsupervised domain adaptation in brain lesion segmentation with adversarial networks', (Springer, 2017, edn.), pp. 597-609
22 Le, M.H., Chen, J., Wang, L., Wang, Z., Liu, W., Cheng, K.-T.T., and Yang, X.: 'Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks', Physics in Medicine & Biology, 2017, 62, (16), pp. 6497
23 Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: 'Going deeper with convolutions', (2015, edn.), pp. 1-9
24 He, K., Zhang, X., Ren, S., and Sun, J.: 'Deep residual learning for image recognition', (2016, edn.), pp. 770-778
25 Yang, X., Liu, C., Wang, Z., Yang, J., Le Min, H., Wang, L., and Cheng, K.-T.T.: 'Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI', Medical image analysis, 2017, 42, pp. 212-227
26 Chen, Q., Hu, S., Long, P., Lu, F., Shi, Y., and Li, Y.: 'A transfer learning approach for malignant prostate lesion detection on multiparametric MRI', Technology in cancer research & treatment, 2019, 18, pp. 1533033819858363
27 Song, Y., Zhang, Y.D., Yan, X., Liu, H., Zhou, M., Hu, B., and Yang, G.: 'Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI', Journal of Magnetic Resonance Imaging, 2018, 48, (6), pp. 1570-1577
28 Kiraly, A.P., Abi Nader, C., Tuysuzoglu, A., Grimm, R., Kiefer, B., El-Zehiry, N., and Kamen, A.: 'Deep convolutional encoder-decoders for prostate cancer detection and classification', (Springer, 2017, edn.), pp. 489-497
29 Mehrtash, A., Sedghi, A., Ghafoorian, M., Taghipour, M., Tempany, C.M., Wells III, W.M., Kapur, T., Mousavi, P., Abolmaesumi, P., and Fedorov, A.: 'Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks', (International Society for Optics and Photonics, 2017, edn.), pp. 101342A
30 Fawcett, T.: 'An introduction to ROC analysis', Pattern recognition letters, 2006, 27, (8), pp. 861-874
31 Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., and Huisman, H.: 'Cancer Imaging Archive Wiki', (2017, edn.), pp.
32 Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., and Huisman, H.: 'Computer-aided detection of prostate cancer in MRI', IEEE transactions on medical imaging, 2014, 33, (5), pp. 1083-1092
33 Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., and Pringle, M.: 'The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository', Journal of digital imaging, 2013, 26, (6), pp. 1045-1057
34 Aldoj, N., Lukas, S., Dewey, M., and Penzkofer, T.: 'Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network', Eur Radiol, 2020, 30, (2), pp. 1243-1253
35 Salman, S., and Liu, X.: 'Overfitting mechanism and avoidance in deep neural networks', arXiv preprint arXiv:1901.06566, 2019

36 Lawrence, S., and Giles, C.L.: 'Overfitting and neural networks: conjugate gradient and backpropagation', (IEEE, 2000, edn.), pp. 114-119
37 Liu, S., Zheng, H., Feng, Y., and Li, W.: 'Prostate cancer diagnosis using deep learning with 3D multiparametric MRI', (International Society for Optics and Photonics, 2017, edn.), pp. 1013428
38 Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W.: 'On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation', PloS one, 2015, 10, (7), pp. e0130140
39 Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A.: 'Learning deep features for discriminative localization', (2016, edn.), pp. 2921-2929

## Statutory Declaration

"I, Nader Aldoj, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic 'A 3D Multi-channel convolutional neural network for classification of prostate cancer using multiparametric MR imaging' | 'Ein 3D-Mehrkanal-Faltungsnetzwerk zur Klassifizierung von Prostatakrebs mithilfe der multiparametrischen MR-Bildgebung', independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me."

Date

Signature

Nader Aldoj

Digital unterschrieben von Nader Aldoj
Datum: 2020.12.01 13:47:11 +01'00'

# Declaration of your own contribution to the top-journal publication for a PhD or MD/PhD degree

Publication:
Aldoj, N., Lukas, S., Dewey, M., Penzkofer, T. (2019). 'Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network'. European Radiology 2019.
DOI 10.1007/s00330-018-5899-8

Nader Aldoj contributed the following to the below listed publication:

Contribution in detail:

- Topic (Aldoj, Penzkofer, Dewey)
- Development of hypotheses, conception, preliminary studies (Penzkofer, Aldoj)
- Software development, method establishment (Aldoj, Lukas)
- Data preparation, processing and analysis (Aldoj)
- Evaluation of the results (qualitative, quantitative, diagnostic, statistical) (Aldoj)
- Evaluation of image quality (Aldoj)
- Discussion of method and results (Penzkofer, Aldoj)
- Manuscript creation, visualizations (all figures and tables) (Aldoj)
- Proofreading (Dewey, Penzkofer, Lukas)
- Presentation (European Congress of Radiology, Vienna 03.2019) (Aldoj)

Marc Dewey

Digital unterschrieben von Marc Dewey
Datum: 2020.12.01 15:42:36 +01'00'

Signature, date and stamp of first supervising university professor / lecturer

Nader Aldoj

Digital unterschrieben von Nader Aldoj
Datum: 2020.12.01 13:46:38 +01'00'

Signature of doctoral candidate

## Extract from the „Journal Summary List“

Journal Data Filtered By: **Selected JCR Year: 2017** Selected Editions: SCIE,SSCI
Selected Categories: **“RADIOLOGY, NUCLEAR MEDICINE and MEDICAL IMAGING”** Selected Category Scheme: WoS
**Gesamtanzahl: 128 Journale**

| Rank | Full Journal Title | Total Cites | Journal Impact Factor | Eigenfactor Score |
|---|---|---|---|---|
| 1 | JACC-Cardiovascular Imaging | 8,104 | 10.247 | 0.026360 |
| 2 | European Heart Journal-Cardiovascular Imaging | 4,630 | 8.336 | 0.020640 |
| 3 | EUROPEAN JOURNAL OF NUCLEAR MEDICINE AND MOLECULAR IMAGING | 14,983 | 7.704 | 0.024870 |
| 4 | RADIOLOGY | 54,109 | 7.469 | 0.063710 |
| 5 | JOURNAL OF NUCLEAR MEDICINE | 27,101 | 7.439 | 0.037560 |
| 6 | CLINICAL NUCLEAR MEDICINE | 4,756 | 6.281 | 0.006950 |
| 7 | INVESTIGATIVE RADIOLOGY | 6,486 | 6.224 | 0.012410 |
| 8 | Circulation-Cardiovascular Imaging | 5,438 | 6.221 | 0.020160 |
| 9 | IEEE TRANSACTIONS ON MEDICAL IMAGING | 17,837 | 6.131 | 0.024200 |
| 10 | ULTRASOUND IN OBSTETRICS & GYNECOLOGY | 12,420 | 5.654 | 0.018820 |
| 11 | INTERNATIONAL JOURNAL OF RADIATION ONCOLOGY BIOLOGY PHYSICS | 46,595 | 5.554 | 0.055060 |
| 12 | JOURNAL OF CARDIOVASCULAR MAGNETIC RESONANCE | 4,918 | 5.457 | 0.013530 |
| 13 | NEUROIMAGE | 92,719 | 5.426 | 0.152610 |
| 14 | MEDICAL IMAGE ANALYSIS | 6,383 | 5.356 | 0.011900 |
| 15 | RADIOTHERAPY AND ONCOLOGY | 17,184 | 4.942 | 0.027840 |
| 16 | HUMAN BRAIN MAPPING | 20,334 | 4.927 | 0.042810 |
| 17 | SEMINARS IN NUCLEAR MEDICINE | 2,285 | 4.558 | 0.002990 |
| 18 | ULTRASCHALL IN DER MEDIZIN | 2,201 | 4.389 | 0.004310 |
| 19 | MAGNETIC RESONANCE IN MEDICINE | 31,440 | 4.082 | 0.034130 |
| **20** | **EUROPEAN RADIOLOGY** | **18,615** | **4.027** | **0.034120** |
| 20 | SEMINARS IN RADIATION ONCOLOGY | 2,480 | 4.027 | 0.003620 |
| 22 | JOURNAL OF NUCLEAR CARDIOLOGY | 3,508 | 3.847 | 0.004120 |
| 23 | AMERICAN JOURNAL OF NEURORADIOLOGY | 22,667 | 3.653 | 0.029840 |
| 24 | JOURNAL OF MAGNETIC RESONANCE IMAGING | 16,398 | 3.612 | 0.027440 |
| 25 | MOLECULAR IMAGING AND BIOLOGY | 2,415 | 3.608 | 0.005480 |

## Original Publication

Aldoj, N., Lukas, S., Dewey, M., Penzkofer, T. (2019). 'Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network'. European Radiology 2019. https://doi.org/10.1007/s00330-019-06417-z

# Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht

# List of publications

Aldoj, N., Lukas, S., Dewey, M., Penzkofer, T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *Eur Radiol* **30,** 1243–1253 (2020).
Impact factor 4.027, https://doi.org/10.1007/s00330-019-06417-z

Asbach, P., Ro, S., Aldoj, N., Snellings, J., Reiter, R., Lenk, J., Köhlitz, T., Haas, M., Guo, J., Hamm, B., Braun, J., Sack, I. In vivo quantification of water diffusion, stiffness and tissue fluidity in benign prostatic hyperplasia and prostate cancer. Investigative Radiology June 2, 2020 - Volume Publish Ahead of Print
Impact factor 6.1, doi: 10.1097/RLI.0000000000000685

Aldoj, N., Biavati, F., Michallek, F., Stober, S., and Dewey, M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. Scientific Reports.
Impact factor 4.12, www.nature.com/articles/s41598-020-71080-0

# Presentations

Nader Aldoj. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network, European Congress of Radiology, Vienna 03.2019.

## Acknowledgment

I would like to thank both the BIOQIC program and its coordination committee, Prof. Ingolf Sack, Dr. Judith Bergs, Prof. Tobias Schäffter, Prof. Sebastian Stober and my direct supervisor Prof. Dr. med. Marc Dewey. I would like to thank the members of the AG Dewey for the personal relationships and the exciting scientific exchange as well as the help with organizational and personal questions. I was supported by Ms. Bettina Herwig for the linguistic examination of the manuscript. Many thanks to my parents and for their always loving support. and finally a special thank you to the German Research Community (GRK2260, BIOQIC) for funding this project.

# Supplementary Material (SM)

## K-trans coefficients:

According to the dataset providers, K-trans images were generated using the parameters of the dynamic contrast enhanced sequence. This is done by a software developed in Radboud University, which fits the MR signal of the enhancement-time curve to an exponential signal intensity model that is described with five parameters: base line signal enhancement, start of the signal enhancement, time to peak (TTP), peak enhancement and wash out.

$$v_e = \frac{P_{tissue}}{P_{plasma}}$$

$$K_{ep} = \frac{1}{\left(TTP_{tissue} - TTP_{plasma}\right)}$$

$$K^{trans} = v_e . K_{ep}$$

$P$ denotes the plateau of the gadolinium concentration, $v_e$ is the extracellular volume's estimate, $K_{ep}$ is the rate between the extracellular extravascular space and the plasma space, and finally $K^{trans}$ is the transfer volume constant [1].

## References:


1 Futterer, J.J., Heijmink, S.W., Scheenen, T.W., Veltman, J., Huisman, H.J., Vos, P., de Kaa, C.A.H.V., Witjes, J.A., Krabbe, P.F., Heerschap, A. and Barentsz, J.O. (2006). Prostate cancer localization with dynamic contrast-enhanced MR imaging and proton MR spectroscopic imaging. Radiology, 241(2), 449-458