



Universiteit  
Leiden  
The Netherlands

## **Radiogenomics analysis linking multiparametric MRI and transcriptomics in prostate cancer**

Fernandes, C.D.; Schaap, A.; Kant, J.; Houdt, P. van; Wijkstra, H.; Bekers, E.; ... ; Turco, S.

### **Citation**

Fernandes, C. D., Schaap, A., Kant, J., Houdt, P. van, Wijkstra, H., Bekers, E., ... Turco, S. (2023). Radiogenomics analysis linking multiparametric MRI and transcriptomics in prostate cancer. *Cancers*, 15(12). doi:10.3390/cancers15123074

Version: Publisher's Version








License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3759982>

**Note:** To cite this publication please use the final published version (if applicable).

## Article

# Radiogenomics Analysis Linking Multiparametric MRI and Transcriptomics in Prostate Cancer

Catarina Dinis Fernandes <sup>1</sup>, Annekoos Schaap <sup>1,†</sup>, Joan Kant <sup>2,†</sup>, Petra van Houdt <sup>3</sup>, Hessel Wijkstra <sup>1,4</sup>, Elise Bekers <sup>5</sup>, Simon Linder <sup>6</sup>, Andries M. Bergman <sup>6,7</sup>, Uulke van der Heide <sup>3</sup>, Massimo Mischi <sup>1</sup>, Wilbert Zwart <sup>2,6</sup>, Federica Eduati <sup>2,8</sup> and Simona Turco <sup>1,\*</sup>

<sup>1</sup> Electrical Engineering Department, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; m.mischi@tue.nl (M.M.)

<sup>2</sup> Biomedical Engineering—Computational Biology Department, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands; w.zwart@nki.nl (W.Z.); f.eduati@tue.nl (F.E.)

<sup>3</sup> Department of Radiation Oncology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

<sup>4</sup> Department of Urology, Amsterdam University Medical Centers, 1100 DD Amsterdam, The Netherlands

<sup>5</sup> Department of Pathology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

<sup>6</sup> Division of Oncogenomics, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

<sup>7</sup> Division of Medical Oncology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

<sup>8</sup> Institute for Complex Molecular Systems, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

\* Correspondence: s.turco@tue.nl

† These authors contributed equally to this work.

**Simple Summary:** Prostate cancer is a global health burden. Multi-parametric magnetic resonance imaging is the recommended imaging modality for diagnosis. The recommended treatment differs based on tumor aggressiveness, typically assessed with the use of invasive techniques such as tumor biopsies. By studying the relationship between imaging characteristics and the genomic information obtained from tumor biopsies, it might be possible to detect aggressive tumor characteristics based solely on imaging, which could eventually be used to non-invasively inform on patient-tailored treatments. In this study, we extracted a large number of imaging features and found significant correlations between them and the aggressiveness of the tumor. We additionally investigated transcriptomic features (i.e., patterns of gene expression) associated with tumor aggressiveness and identified significant correlations with perfusion-related image features, highlighting a link between what is visible on the diagnostic images and the underlying genomic information of the tumors.

**Abstract:** Prostate cancer (PCa) is a highly prevalent cancer type with a heterogeneous prognosis. An accurate assessment of tumor aggressiveness can pave the way for tailored treatment strategies, potentially leading to better outcomes. While tumor aggressiveness is typically assessed based on invasive methods (e.g., biopsy), radiogenomics, combining diagnostic imaging with genomic information can help uncover aggressive (imaging) phenotypes, which in turn can provide non-invasive advice on individualized treatment regimens. In this study, we carried out a parallel analysis on both imaging and transcriptomics data in order to identify features associated with clinically significant PCa (defined as an ISUP grade  $\geq 3$ ), subsequently evaluating the correlation between them. Textural imaging features were extracted from multi-parametric MRI sequences (T2W, DWI, and DCE) and combined with DCE-derived parametric pharmacokinetic maps obtained using magnetic resonance dispersion imaging (MRDI). A transcriptomic analysis was performed to derive functional features on transcription factors (TFs), and pathway activity from RNA sequencing data, here referred to as transcriptomic features. For both the imaging and transcriptomic features, different machine learning models were separately trained and optimized to classify tumors in either clinically insignificant or significant PCa. These models were validated in an independent cohort and model performance was used to isolate a subset of relevant imaging and transcriptomic features to be further investigated. A final set of 31 imaging features was correlated to 33 transcriptomic features obtained on the same tumors. Five significant correlations ( $p < 0.05$ ) were found, of which,



**Citation:** Dinis Fernandes, C.; Schaap, A.; Kant, J.; van Houdt, P.; Wijkstra, H.; Bekers, E.; Linder, S.; Bergman, A.M.; van der Heide, U.; Mischi, M.; et al. Radiogenomics Analysis Linking Multiparametric MRI and Transcriptomics in Prostate Cancer. *Cancers* **2023**, *15*, 3074. <https://doi.org/10.3390/cancers15123074>

Academic Editor: Mary Frances McMullin

Received: 19 April 2023

Revised: 11 May 2023

Accepted: 16 May 2023

Published: 6 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

three had moderate strength ( $|r| \geq 0.5$ ). The strongest significant correlations were seen between a perfusion-based imaging feature—MRDI *A* median—and the activities of the TFs STAT6 (−0.64) and TFAP2A (−0.50). A higher-order T2W textural feature was also significantly correlated to the activity of the TF STAT6 (−0.58). STAT6 plays an important role in controlling cell proliferation and migration. Loss of the AP2alpha protein expression, quantified by TFAP2A, has been strongly associated with aggressiveness and progression in PCa. According to our findings, a combination of texture features extracted from T2W and DCE, as well as perfusion-based pharmacokinetic features, can be considered for the prediction of clinically significant PCa, with the pharmacokinetic MRDI *A* feature being the most correlated with the underlying transcriptomic information. These results highlight a link between quantitative imaging features and the underlying transcriptomic landscape of prostate tumors.

**Keywords:** magnetic resonance imaging; prostate cancer; radiogenomics; machine learning

## 1. Introduction

Prostate cancer (PCa) has the highest incidence among men worldwide, accounting for one-third of all cancer diagnoses in males [1]. In 2020, with 375,000 deaths worldwide, PCa was the fifth leading cause of cancer-related death among men, taking the lead in countries located in sub-Saharan Africa, the Caribbean, and Central and South America [2]. In the traditional diagnostic pathway of PCa, patients with suspicious findings on serum prostate-specific antigen (PSA) testing and/or digital rectal examination (DRE) are typically referred for biopsy, with or without prior imaging [3]. The risk for biochemical recurrence is then calculated based on PSA values at diagnosis, the Gleason score at biopsy, and the clinical stage [3]. While active surveillance is recommended for low-risk disease, radical treatment (radical prostatectomy or radiotherapy) is offered to intermediate-risk patients, and multimodal therapeutic approaches, combining active treatment and systemic therapies, should be considered for advanced, high-risk PCa [4]. This pathway has been shown to lead to both underdiagnosis and misdiagnosis of clinically significant PCa, as well as overdiagnosis and overtreatment of insignificant PCa [4–6].

Spurring from a need for more accurate diagnostic and prognostic markers, genomic-based tests have been proposed and validated [7,8]. However, the costs and incomplete evidence for long-term outcomes have hampered the clinical implementation of genomic tests [9]; currently, they are only advised when risk upgrading may change clinical management [9] and for advanced PCa [10]. Additionally, as genomic tests are performed on biopsied samples, they bear the same risks (infection and bleeding) and discomfort associated with biopsies [11].

Multi-parametric magnetic resonance (mpMRI), combining anatomical T2-weighted imaging, with functional diffusion-weighted imaging (DWI) and dynamic-contrast enhanced (DCE) imaging [5,12], is an established non-invasive tool for PCa imaging with the potential to improve PCa diagnosis and management [4–6,13]. According to the latest updates from the American and European guidelines [3,4,9,14], mpMRI is recommended for early detection in patients with suspicious PSA and/or DRE findings, for targeting biopsies, for augmenting risk-stratification in active surveillance patients, and for local tumor staging. However, despite improvements achieved through the updated prostate Prostate Imaging–Reporting and Data System (PI-RADS) scoring [15], low specificity and inter-observer variability remain concerns [4,5,14].

In the past decades, advanced computational approaches for quantitative mpMRI analyses have shown promise for (semi)automatic extraction of imaging features, possibly improving diagnostic accuracy, while mitigating the unavoidable variability inherent in human-based interpretation and scoring systems [5,12,16,17]. While the calculation of the apparent diffusion coefficient (ADC) from DWI [18] and the pharmacokinetic analysis of the DCE sequence [19,20] are well-established quantification methods for prostate mpMRI,

radiomic analysis has recently emerged for the high-throughput extraction of imaging features, ranging from size, shape, and volume, to statistical properties of the image intensity, and more complex fractal, texture, and transform-based features [21–24]. Currently, radiomic features typically require the identification and manual delineation of a region of interest, such as a tumor lesion. From this region, one value of each feature is extracted, thus losing the granularity available in ADC and pharmacokinetic maps, in which, instead, the parameters are extracted at each voxel.

Aligned with the current paradigm shift towards precision oncology, the concept of radiogenomics has more recently emerged, aimed at linking radiomic features with the underlying molecular and genomic characteristics of tumor tissues and micro-environments [21,24,25]. The overarching goal is to identify radiomic features that correlate with and complement genomic-based markers for improved diagnosis, risk stratification, and therapeutic decision-making. Although the word “genomics” technically refers to the presence of genes, in the context of radiogenomics, it is also often used to refer to gene expression levels (transcriptomics).

In the prostate, initial studies on radiogenomics focused on the association of the PCA3 score, a urine marker for aggressive prostate cancer, with a qualitative assessment of mpMRI by PI-RADS or Likert scoring [26–28], with conflicting results. While a significant association was found between the PCA3 score and PI-RADS [28], PCA3 was not associated with cancer detection for high-suspicion mpMRI cases [27]. Several studies investigated the association between mpMRI findings (e.g., tumor visibility and PI-RADS) with transcriptomics-based scores and molecular biomarkers from biopsied cancer tissue, including Prolaris, Decipher, the genomic prostate score, and expression of phosphatase and tensin homolog (PTEN) [21,29–32]. Although transcriptomics scores were generally predictive of adverse pathology results (the Gleason score) [27,29,31,32] and an association was found between PI-RADS and genomic scores [30,31], a wide range of genomic-based score values was observed across all PI-RADS categories [29,30]. Beyond a qualitative assessment of mpMRI, a few studies extracted quantitative features by e.g., radiomics [33–38]. Focusing only on PTEN expression, McCann et al. [33] found a significant correlation with  $k_{ep}$ , a pharmacokinetic parameter extracted from DCE, which is related to vascular permeability. However, this result is in contrast with a later study, in which PTEN expression only correlated with mean ADC values [36]. Investigating tumor hypoxia, Sun et al. [38] found a set of texture features extracted from T2W imaging to be significantly associated with hypoxia-related genes. While previous studies only focused on one or a few genes, broader genes and microRNA expression analyses were performed in [35,37]. Both studies demonstrated that radiomics features are correlated with gene signatures related to immune response activation, cycle cell processes, and DNA repair [35]; a radiomic-based classifier was able to predict Gleason scores and Decipher scores with moderate accuracy [35]. With few exceptions [33,36], however, the DCE series is not included in the quantitative mpMRI analysis; this is possibly due to the hurdles of PK analysis by the conventional Tofts model [19], which requires the separate estimation of an arterial input function. Moreover, only a subset of intensity-based and texture-based features was extracted from the whole tumor ROI.

In this study, we investigate the relationship between imaging and transcriptomic features in aggressive PCa. To this end, we first identify the imaging and transcriptomic features that are most predictive of aggressive PCa by machine learning and then analyze their correlation. In contrast to previous studies, we investigate a larger set of texture features on all mpMRI imaging sequences and we adopt a moving-window strategy to extract features at each imaging voxel, allowing the investigation of the feature spatial distribution and heterogeneity in the tumor ROI. Additionally, we perform magnetic resonance dispersion imaging (MRDI) on the DCE sequence for automatic estimation of PK parameters related to microvascular perfusion, architecture, and permeability, without the need for an arterial input function [20,39].

## 2. Materials and Methods

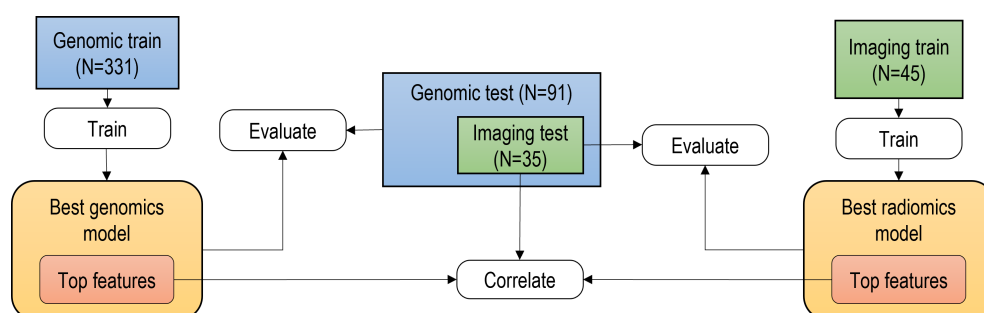
### 2.1. Data Collection

In this work, the genomic model was based on transcriptomic features, while the radiomics model was based on imaging features extracted from mpMRI. As described in Figure 1, three independent datasets were used in this study. For both genomics and imaging, separate training and test datasets were used; the training dataset was used for the optimization of machine learning model(s) and feature selection, while the test dataset was used for performance evaluation and for investigating the correlation between the most relevant imaging and transcriptomic features.

The genomics training dataset consisted of prostate adenocarcinoma patients from the TCGA dataset. Bulk RNA-seq data were obtained through the online tool from the Broad Institute (<https://gdac.broadinstitute.org/>, version released 28 January 2016), and clinical information was downloaded from the GDC Data Portal (<https://portal.gdc.cancer.gov/>, accessed on 18 September 2022). We used a subset of 331 patients passing quality control and with revised clinical data, as described in [40].

The imaging training dataset was retrospectively collected from 3 different Dutch institutes (Amsterdam University Medical Center, The Netherlands Cancer Institute, and Radboud University Medical Center), as part of the Prostate Cancer Molecular Medicine study. Patients were eligible for inclusion if they underwent radical prostatectomy due to biopsy-proven PCa and if a preoperative mpMRI, including T2W, DWI, and DCE sequences, was available. Reasons for exclusion were: the presence of (movement) artifacts in one of the sequences, the temporal resolution of the DCE sequence was not sufficiently high for the MRDI analysis, and missing histopathology analysis. According to the inclusion criteria, a total of 45 patients were included.

For both imaging and genomics, the test dataset consisted of PCa patients treated at The Netherlands Cancer Institute (Amsterdam, the Netherlands). The genomics test dataset included 91 patients, from whom bulk RNA-seq was taken from a single identified lesion on the biopsy sample. In the sporadic case when multiple lesions were visible, the RNA sample was taken from the index lesion. A subset of patients was also imaged with mpMRI and biopsied prior to radical prostatectomy, which constituted the imaging test dataset. The inclusion criteria for the imaging test dataset were the same as for the imaging training dataset, with the addition of the RNA sample being available, resulting in 35 patients being included.



**Figure 1.** General framework and overview of the datasets used for the radiogenomics analysis. In this work, the genomic model was based on transcriptomic features, while the radiomics model was based on features extracted from mpMRI.

The histopathological analysis of the prostate after radical prostatectomy was available for each patient included in the study. The pathologically derived ISUP grade [41] was then used to stratify patients into clinically insignificant ( $ISUP \leq 2$ ) and significant ( $ISUP \geq 3$ ) PCa. An overview of the distributions of ISUP grades for all patients included in this study is provided in Table 1.

In the imaging training dataset, prostate specimens were first fixated in formalin and then cut into 4 mm slices and photographed. The microscopic analysis was then performed

by a pathologist to mark cancer regions on the basis of cellular differentiation. In the test dataset, the surgically removed prostate was stained with hematoxylin and eosin (H&E), and a pathologist delineated the tumor in the available pathology slides.

**Table 1.** Overview of ISUP grades (index lesion) given in terms of the number of patients and respective percentages in each dataset (in parenthesis).

	Transcriptomics		Imaging	
	Training	Test	Training	Test
ISUP 1	64 (19.3%)	5 (5.5%)	11 (24.4%)	1 (2.9%)
ISUP 2	102 (30.8%)	29 (31.9%)	19 (42.2%)	10 (28.6%)
ISUP 3	78 (23.6%)	27 (29.7%)	9 (20.0%)	10 (28.6%)
ISUP 4	44 (13.3%)	15 (16.5%)	3 (6.7%)	8 (22.9%)
ISUP 5	43 (13.0%)	15 (16.5%)	3 (6.7%)	6 (17.1%)

### 2.1.1. Imaging Data Pre-Processing

The MRI acquisition settings are summarized in the supplementary materials, Tables S1 and S2 for the training and test datasets, respectively. While for the training dataset, ADC maps automatically generated by the scanner were available. For the test dataset, the ADC maps were often missing and/or of poor quality; therefore, new maps were generated for the whole test dataset. A minimum of two  $b$  values between 100 and 1000  $s/mm^2$  were used to calculate the maps. As described in the PI-RADS guidelines [42], the ADC maps were generated using the conventional monoexponential model, where the signal intensity on DWI has a log-linear relationship with the  $b$  values. The best-fitting linear equation was found using a least squares approximation, with the slope yielding the ADC value.

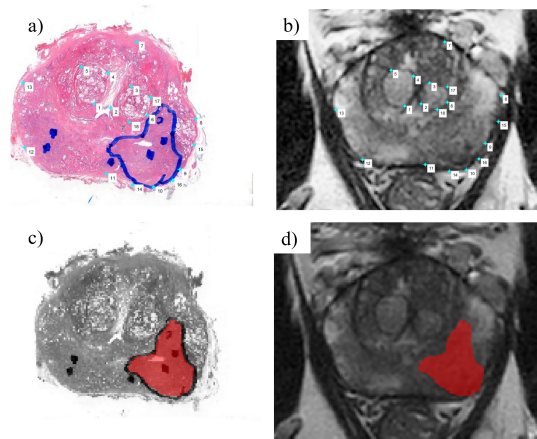
Prostate zonal segmentation was performed on all patients using a modified multi-channel U-net neural network to extract the contours of the whole prostate, the central gland, and the peripheral zone [43]. The obtained contours were reviewed and adjusted manually when necessary. For both datasets, the three imaging sequences were resampled onto the T2W imaging grid for each patient.

For both training and test datasets, tumor regions of interest (ROIs) were defined based on the histopathology contours. In the training dataset, tumor delineations were available for each histopathology slide, which were used to guide visual matching and manual delineation of a tumor ROI on the MRI scans using ITK snap [44]. In the test dataset, one to two H&E-stained slides were available per patient. These were then scanned and deformably registered to the T2W grid. Firstly, the T2W slice that best matched the H&E stained slide was identified, and then anatomical landmarks were marked on both images (Figure 2a,b). The H&E samples were deformably registered to the grid of T2W MRI by using a coherent point drift algorithm, based on the manually selected matching landmarks on both images. The transformed H&E slide (Figure 2c) was then used to draw and propagate the tumor delineations onto the T2W slice (Figure 2d). The tumor ROI was delineated using ITK snap [44].

In order to account for registration errors in defining the tumor ROIs, the registration error was derived per patient by measuring the distance between matching visual landmarks on both images. The tumor ROIs were then eroded with an error margin of half the average registration error. The size of the error was determined to be 1.19 mm. Therefore, the tumor contours were eroded with a margin of 0.594 mm (approximately 2 pixels in the T2W grid).

For training, the full extent of the identified tumor lesions—often comprising several MRI slices after matching to the corresponding pathology slides—were included. In the sporadic case when there were multiple lesions, the GS of the index lesion was assigned. In the test dataset, for each patient, one H&E slide was successfully registered to a corresponding MRI slice. Additionally, as the transcriptomics information was only available for the index lesion, the analysis of the test dataset focused on the index tumor ROI as identified

in the single MRI slice with histopathology matching. This resulted in a total of 218 and 35 lesion ROIs for model training and testing, respectively.



**Figure 2.** Illustration of the registration process and propagation of the tumor contour from the H&E slide to the matching T2W slice. (a) H&E-stained pathology slide with reference landmarks; (b) the matched T2W slice with reference landmarks; (c) deformably registered H&E slide; (d) H&E tumor contour propagated to the T2W slice.

## 2.2. Transcriptomics Analysis Pipeline

### 2.2.1. Feature Extraction

To summarize the transcript expression levels, the counts and transcript per million (TPM) were computed from the bulk RNA sequencing data by using the first two steps of the pipeline from *quantiseq* [45], where RNA sequencing reads were pre-processed with *Trimmomatic* [46]; consequently, the raw counts and TPM were quantified with *Kallisto* [47]. The RNA sequencing data were processed using the R package *EaSeR* [48], to derive quantitative features describing different facets of the tumor microenvironment, such as the activity of pathways and transcription factors (TFs). In particular, we made use of the *PROGENy* tool [49] to compute the activity of 14 intracellular signaling pathways based on the expressions of genes that are differentially regulated during pathway-specific perturbation experiments. The *DoRothEA* tool [50] was used to infer the activity of 118 TFs based on the expressions of their target genes. In addition, the gene signature scores for *PORTOS* and *Decipher* were derived. *PORTOS* [51] and *Decipher* [52] are two well-established PCa signatures, the first is used to predict the response to postoperative radiotherapy and the latter to predict the response to radical prostatectomy. The model for the *PORTOS* score was developed for microarray data, and the model for *Decipher* was proprietary, as a result, no scores could be computed. However, we explored three alternative methods to compute scores based on the expressions of the signature genes: *ssGSEA* [53], geometric mean [54], and PCA [55]. For the PCA method, we use the values from the first principal component as the scores. In this way, 6 signature scores were obtained (3 from *PORTOS* and 3 from *Decipher*), leading to a total of 138 transcriptomic features. Table S3, included in the supplementary materials, lists the included features.

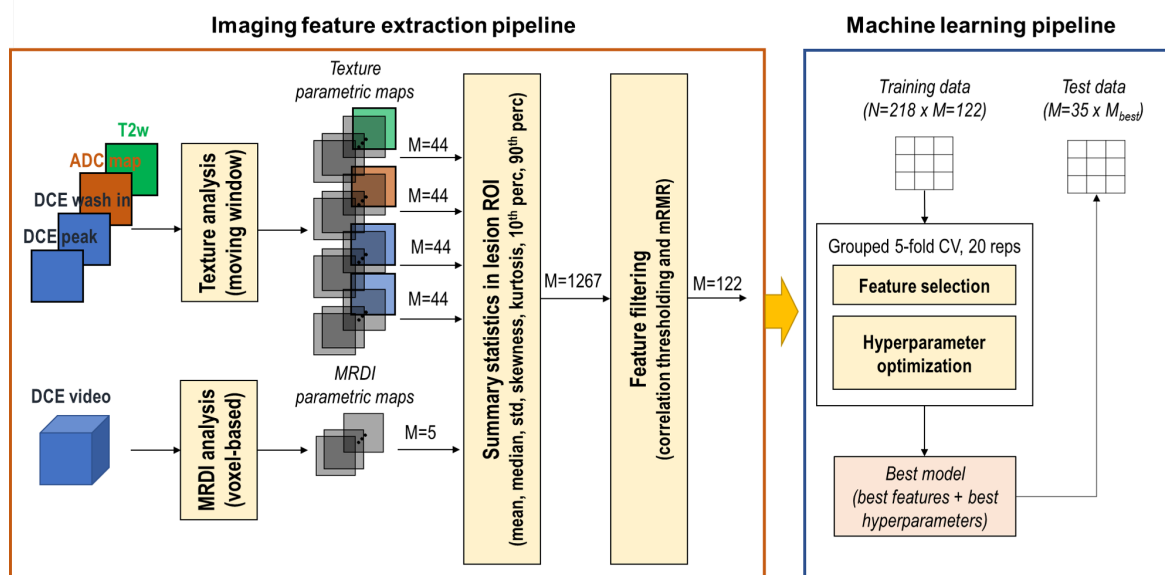
### 2.2.2. Machine Learning for Feature Selection

A regularized logistic regression model was optimized on the transcriptomic training dataset by a nested stratified cross-validation approach with 25 repeats, 10 outer loops, and 5 inner loops. In the nested approach, the inner loop is used to optimize the hyperparameters, while the outer loop is used for performance evaluation and model selection. In both loops, folds were stratified to ensure that the folds were balanced between the two classes (insignificant vs. significant PCa), and features were standardized in each outer loop. The models were optimized by minimizing the log loss and penalized with elastic net regularization. Features were ranked by importance by averaging the weight

coefficients obtained for each feature at each repetition of the cross-validation procedure and converting them to odds ratios (ORs) using  $OR = e^w$ , with  $w$  being the coefficient of a feature. The OR quantifies the relative odds of having clinically significant PCa given the value of a feature, where an  $OR > 1$  means that a higher value for this feature is associated with higher odds of clinically significant PCa and the opposite with an  $OR < 1$  [56]. Features for which the OR was very close to 1, with a margin of 0.1 ( $|OR - 1| < 0.1$ ), were excluded from further investigation. This procedure was done separately to perform feature ranking and selection for the transcription factors and pathway activity. For PORTOS/Decipher signature scores, all six features were further investigated.

### 2.3. Image Analysis Pipeline

An overview of the imaging feature extraction and machine learning pipelines is presented in Figure 3. For each imaging slice where a lesion is delineated, texture analysis by a moving window is applied to T2W, ADC, DCE wash-in, and peak images. In parallel, the MRDI analysis is applied to the full DCE-MRI video to obtain MRDI parametric maps. Summary statistics are calculated from the voxels within the lesion's ROI. Dimensionality reduction is then applied by correlation thresholding and maximum relevance minimum redundancy (mRMR) feature selection [57]. The remaining features were used as input to the machine learning pipeline. For each tested classifier, 20 repetitions of the grouped 5-fold cross-validation were performed on the training data to find the best model, consisting of the best subset of features and the best hyperparameters. The performance of the best model was then evaluated on the test data. All the steps of the pipeline are detailed hereafter.



**Figure 3.** Overview of the imaging feature extraction and machine learning pipelines.  $N$  represents the number of samples,  $M$  represents the number of features.

#### 2.3.1. Feature Extraction

**MRDI analysis.** A pharmacokinetic analysis approach, denoted as magnetic resonance dispersion imaging (MRDI) [20,39], was performed on the DCE-MRI time series in order to derive parameters describing microvascular perfusion, architecture, and permeability. The MRDI analysis consists of a two-compartment model, with one compartment accounting for the intravascular contrast transport, and one compartment accounting for extravasation. While extravasation is modeled as a mono-exponential decay, a convective dispersion model is adopted to describe the contrast agent transport within the (micro)vascular network. A local solution of this model, the modified local density random walk, provides the intravascular contrast concentration,  $C_i(t)$ . As described in [20,39], this

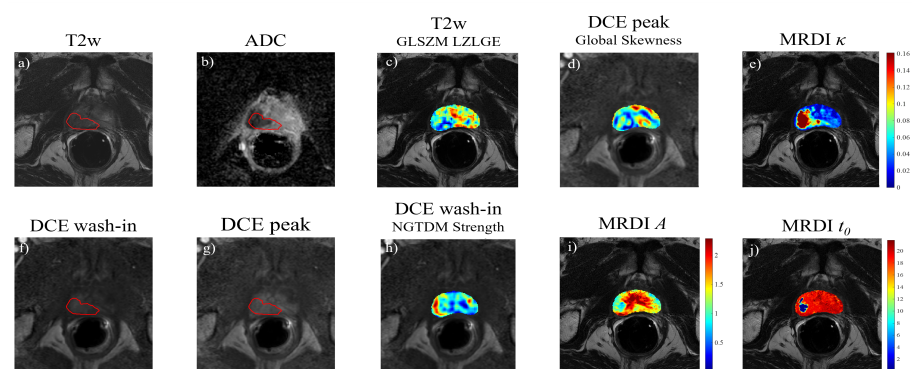
model circumvents the need for arterial input function estimation, and a reduced dispersion model can be used to calculate the PK parameters:

$$C(t) = A \sqrt{\frac{\kappa}{2\pi(t-t_0)}} e^{-\frac{\kappa(t-t_0-\mu)^2}{2(t-t_0)}} * e^{-k_{ep}t}, \quad (1)$$

where  $A = \alpha K^{trans}$ , with  $\alpha$  equal to the time integral of the intravascular concentration and  $K^{trans}$  the forward volume transfer constant;  $\mu$  is the intravascular mean transit time;  $t_0$  is the theoretical injection time; and  $\kappa$  is the dispersion parameter. As a separate estimation of  $K^{trans}$  is not possible, the parameter  $A$  is in turn investigated. Equation (1) was used to fit the time-intensity curves obtained at each voxel of the DCE sequence, resulting in the PK parameters  $A$ ,  $\mu$ ,  $t_0$ ,  $k_{ep}$ , and  $\kappa$ . Figure 4 shows an example of the extracted  $A$ ,  $t_0$ , and  $\kappa$  maps for one patient. The aforementioned analysis was performed using MATLAB R2022a.

**Texture analysis.** While the texture analysis of T2W and DWI is straightforward, due to the dynamic nature of the DCE sequence, suitable temporal frames first need to be selected. As malignant lesions have been associated with distinguishable enhancement patterns at wash-in and higher peak intensities [58], here, the wash-in and peak frames were isolated and used as input for texture feature extraction. The time frames were chosen by analyzing the average time-intensity curve within the prostate region in a single slice located in the middle of the gland. The frame with the maximum mean intensity was defined as the peak frame. As for the wash-in, this was defined as the frame for which the time-intensity curve displayed the highest positive gradient. The candidate wash-in and peak frames were visually inspected and, when necessary, a more representative frame was chosen.

Using the Vallières et al. Matlab toolbox [59], texture feature maps were extracted from T2W, ADC, DCE peak, and DCE wash-in frames (Figure 4). The features were extracted using a moving window of size of  $15 \times 15$  pixels, moved with a stride of 3 pixels. The window size and stride were chosen empirically to prevent the loss of local characteristics of the features while ensuring a meaningful distribution of the intensity values in the window. Compared to the traditional ROI-based feature extraction, this approach allows for the extraction of texture features at each voxel, thus obtaining parametric texture maps (see example in Figure 4). In each window, the intensity was quantized to 64 grayscale levels using equal probability quantization. A total of 43 texture features were extracted for each image, as summarized in Table 2, for a total of 172 texture maps.



**Figure 4.** Example of mpMRI images (a) T2W, (b) ADC, (f) DCE wash-in frame, and (g) DCE peak frame), with corresponding texture feature maps (c) T2W GLSZM LZLGE, (d) DCE peak global skewness, and (h) DCE wash-in NGTDM strength and MRDI parametric maps, (e)  $\kappa$ , (i)  $A$ , and (j)  $t_0$ . The tumor contours are displayed in red.

**Table 2.** Overview of extracted texture features.

Texture Type	Features	Number of Features
Global	Variance, skewness, kurtosis	3
Gray-level co-occurrence matrix (GLCM)	Energy, contrast, correlation, homogeneity, variance, Sum average, entropy, dissimilarity, autocorrelation	9
Gray-level run-length matrix (GLRLM)	Short-run emphasis (SRE), long-run emphasis (LRE), gray-level non-uniformity (GLN), run-length non-uniformity (RLN), run percentage (RP), low gray-level run emphasis (LGRE), high gray-level run emphasis (HGRE), short-run low gray-level emphasis (SRLGE), short-run high gray-level emphasis (SRHGE), long-run low gray-level emphasis (LRLGE), long-run high gray-level emphasis (LRHGE), gray-level variance (GLV), run-length variance (RLV)	13
Gray-level size zone matrix (GLSZM)	Small zone emphasis (SZE), large zone emphasis (LZE), gray-level non-uniformity (GLN), zone-size non-uniformity (ZSN), zone percentage (ZP), low gray-level zone emphasis (LGZE), high gray-level zone emphasis (HGZE), small zone low gray-level emphasis (SZLGE), small zone high gray-level emphasis (SZHGE), large zone low gray-level emphasis (LZLGE), large zone high gray-level emphasis (LZHGE), gray-level variance (GLV), zone-size variance (ZSV)	13
Neighborhood gray-tone difference matrix (NGTDM)	Coarseness, contrast, busines, Complexity, strength	5

### 2.3.2. Summary Statistics and Feature Filtering

In total, 181 parametric maps were obtained, including 4 raw grayscale images (T2W, DWI, DCE\_wash\_in, DCE\_peak), 43 texture maps obtained for each of the 4 raw images, and 5 MRDI maps. In order to capture the feature distribution in the lesion, the following summary statistics were calculated in the lesion ROI for each feature map: median, 10th and 90th percentiles, standard deviation (std), mean, skewness, and kurtosis. This resulted in a total of 1267 features. Redundant features were removed by correlation analysis, in which for each pair of features with a correlation larger or equal to 0.9, only the feature with the largest correlation to the outcome variable was kept. The principal component analysis (PCA) was then used to inform on the number  $m$  of components necessary to explain 95% of the variance. The features were ranked using mRMR [57] and the top  $m$  features were selected as input to the machine learning pipeline, described hereafter.

### 2.3.3. Machine Learning Pipeline

An overview of the machine learning pipeline is schematically shown in Figure 3. After dimensionality reduction, the remaining  $m$  features were given as input to a grouped stratified 5-fold cross-validation procedure, which ensured that lesions from the same patient were always grouped in the same fold; in this way, data from the same patient can never be in both the training and validation folds. Additionally, folds were stratified to preserve class proportions. The procedure was repeated 20 times with different splits of the patients in the folds.

At each iteration of the cross-validation loop, feature selection was performed to determine both the best number and subset of features, and the hyperparameters of three machine learning models—logistic regression (LR), support vector machine (SVM), and k-nearest neighbors (kNN)—were optimized. The LR, KNN, and SVM models were chosen as they are widely used, simple to implement, and computationally efficient. The chosen optimization criterion was the area under the ROC curve (AUC), which balances between sensitivity (SENS) and specificity (SPEC). To cope with class imbalance, the “class weight” hyperparameter was set to “balanced” for the LR and SVM; however, this was not possible for the kNN. For all classifiers, feature selection was performed by recursive feature selection, using a basic logistic regression as a classifier. Additionally, to investigate the importance of each feature in the prediction, the permutation feature importance (PFI)

was calculated for each optimized classifier. Briefly, the PFI is calculated as the decrease in the model score after randomly permuting the values of one feature, so that the feature loses any relationship with the class. In this way, it is possible to individually ascertain the importance of each feature in a model-agnostic fashion. The calculation was repeated with 10 random permutations, and the mean PFI was then normalized so that the most important feature has a mean PFI equal to 1.

Finally, the optimized models were validated on the test dataset, for which the PFI was also calculated. The performances of the models were evaluated on both training and test datasets by calculating sensitivity, specificity, accuracy (ACC), balanced accuracy (bACC), and AUC.

All steps of the machine learning pipeline were performed using scikit-learn [60].

#### 2.4. Correlation with Transcriptomic Features

The selected transcriptomic features were correlated with the most relevant imaging features, i.e., the selected features for the best classifier. The correlations were corrected using false discovery rate correction for multiple testing by the Benjamini–Hochberg method [61], with a family-wise error rate set to 0.2. The correlations were clustered using unsupervised hierarchical clustering with Euclidean distance as a similarity metric with the R package ComplexHeatmap [62]. Only significant correlations ( $\alpha = 0.05$ ) were further investigated.

### 3. Results

#### 3.1. Patient Stratification

According to the ISUP-based criterion described in Section 2.1, the transcriptomic training data consisted of 165 clinically insignificant and 166 clinically significant PCa, while in the test dataset, these were 57 and 34, respectively. For the imaging data, the training dataset included 30 clinically insignificant PCa patients, yielding 140 ROIs, and 15 clinically significant PCa patients, yielding 78 ROIs. This resulted in a total of 218 ROIs used to train and optimize the model parameters. For the test imaging data, there were 11 clinically insignificant and 24 clinically significant PCa patients, each providing one ROI, as explained in Section 2.1.1.

#### 3.2. Dimensionality Reduction and Feature Selection

First, we performed knowledge-based dimensionality reduction on the transcriptomics data by deriving 132 functional features describing pathways and TF activity from the expressions of more than 20,000 genes, as described in Section 2.2.1. Then the transcriptomic features were selected according to the OR criterion described in Section 2.2.2, resulting in 10 pathways and 17 transcription factors. Together with the 6 PORTOS/Decipher signature scores, a total of 33 transcriptomic features were input to the correlation analysis—Table 3. For the imaging features, removing highly correlated features reduced the feature set from 1267 to 609. A total of 122 principal components was necessary to explain 95% variance within the imaging training dataset, and this number was used to empirically define the subset of features to select for further analysis. Thus, the top-ranking mRMR 122 features were used to further optimize and train the models.

**Table 3.** Description of the 33 selected transcriptomic features, consisting of 17 transcription factors, 10 signaling pathways and 6 gene signature scores.

Type of Transcriptomic Feature	Acronym	Full Name
Transcription factors	REST	RE1 silencing transcription factor
	FOXA2	Forkhead box protein A2
	TFDP1	Transcription factor Dp-1

Table 3. Cont.

Type of Transcriptomic Feature	Acronym	Full Name
	ERG	Erythroblast transformation specific (ETS) transcription factor ERG
	SOX9	SRY-box transcription factor 9
	TFAP2A	Transcription factor AP-2 alpha
	TFAP2C	Transcription factor AP-2 gamma
	HNF4A	Hepatocyte nuclear factor 4 alpha
	EPAS1	Endothelial PAS domain-containing protein 1
	E2F1	E2F transcription factor 1
	E2F2	E2F transcription factor 2
	E2F3	E2F transcription factor 3
	E2F4	E2F transcription factor 4
	BACH1	BTB and CNC homology 1
	STAT6	Signal transducer and activator of transcription 6
	AR	Androgen receptor
	NR2F2	Nuclear receptor subfamily 2 group F member 2
Pathways *	EGFR	Epidermal growth factor receptor—regulates growth, survival, migration, apoptosis, proliferation, and differentiation in mammalian cells.
	PI3K	Phosphatidylinositol 3-kinase—promotes growth and proliferation.
	Androgen	Involved in the growth and development of the male reproductive organs.
	Estrogen	Promotes the growth and development of the female reproductive organs.
	JAK-STAT	Janus kinases (JAKs), signal transducer and activator of transcription proteins (STATs)—involved in immunity, cell division, cell death, and tumor formation.
	MAPK	Mitogen-activated protein kinases (MAPKs)—integrates external signals and promotes cell growth and proliferation.
	TGFb	Transforming growth factor beta—involved in the development, homeostasis, and repair of most tissues.
	Trail	TNF-related apoptosis-inducing ligand—induces apoptosis.
	VEGF	Vascular endothelial growth factor—mediates angiogenesis, vascular permeability, and cell migration.
	WNT	Created from the names wingless and Int-1—regulates organ morphogenesis during development and tissue repair.
Signatures	Decipher geometric mean	Decipher signature, derived using the geometric mean
	Decipher PCA	Decipher signature, derived using PCA
	Decipher ssGSEA	Decipher signature, derived using single-sample gene set enrichment analysis
	PORTOS geometric mean	PORTOS signature, derived using the geometric mean
	PORTOS PCA	PORTOS signature, derived using PCA
	PORTOS ssGSEA	PORTOS signature, derived using single-sample gene set enrichment analysis

\* The description provided for each pathway was derived from the PROGENy [49] pathway signatures GitHub repository.

### 3.3. Imaging Classifiers

For each imaging classifier, the performance was evaluated at each repetition of the k-fold cross-validation framework to find the optimal number of features and the best hyper-

parameters. These are presented for each model in Table 4. The performances of the imaging classifiers on both training and test datasets are presented in Table 5. Both LR and SVM achieve good performances on the training set (AUC = 0.82 and AUC = 0.83, respectively), and moderate performances on the test dataset (AUC = 0.74 and AUC = 0.72, respectively).

**Table 4.** Overview of the optimal hyperparameters and optimal number of features for each classifier.

Classifier	Hyperparameter Description	Optimal Value
LR	Solver	Liblinear
	Penalty	l2
	C	0.02
	$N_0$	31
KNN	Weights	Uniform
	Metric	Minkowski
	K	28
	$N_0$	30
SVM	Kernel	Sigmoid
	$\gamma$	0.01
	C	0.95
	$N_0$	27

$N_0$ —Optimal number of features; C—regularization parameter; K—number of neighbors;  $\gamma$ —Kernel coefficient.

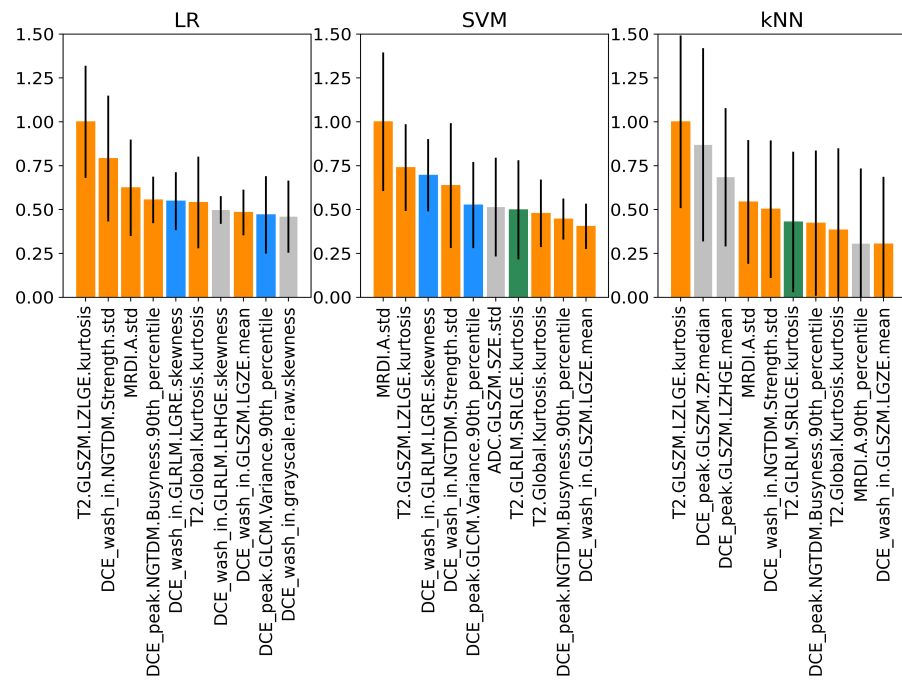
**Table 5.** Performances of the three imaging classifiers on the imaging training and test datasets. The displayed values for the training dataset are the mean and standard deviations (in parenthesis) over the cross-validation scheme.

	Train			Test		
	LR	KNN	SVM	LR	KNN	SVM
ACC	0.72 (0.07)	0.67 (0.04)	0.75 (0.08)	0.71	0.34	0.69
bACC	0.72 (0.08)	0.55 (0.05)	0.74 (0.08)	0.74	0.52	0.72
Sensitivity	0.72 (0.14)	0.15 (0.11)	0.72 (0.16)	0.67	0.04	0.63
Specificity	0.72 (0.09)	0.96 (0.05)	0.76 (0.09)	0.82	1.00	0.82
AUC	0.82 (0.08)	0.72 (0.08)	0.83 (0.08)	0.74	0.52	0.72

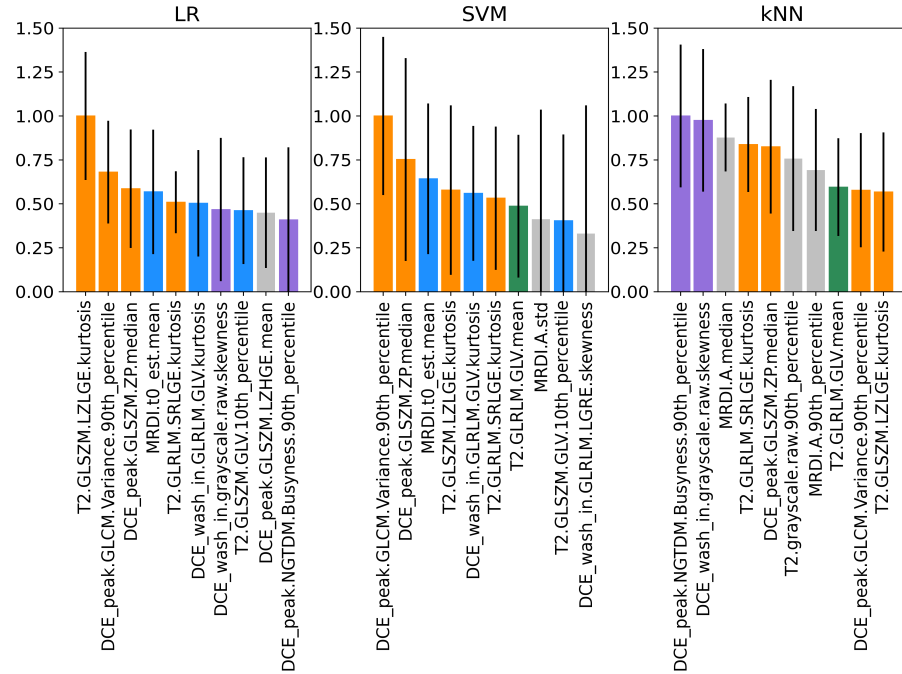
ACC = accuracy; bACC = balanced accuracy; AUC = area under the curve.

### 3.4. Permutation Feature Importance of Imaging Features

The normalized PFIs of the top 10 features obtained on the training dataset for all imaging classifiers are shown in Figure 5. As shown by the coloring, most features were shared by the classifiers. Interestingly, LR and SVM share three of their top four features, of which two are also shared with the kNN. These are the standard deviation of MRDI parameter  $A$  and the kurtosis of the texture feature GLSZM-LZLGE extracted from T2W. Similar findings are observed in the test set (Figure 6), where several features are common among all classifiers, and the top four features are the same for LR and kNN. Notably, for LR and SVM, the kurtosis of the texture feature GLSZM-LZLGE from T2W and the 90th percentile of the texture feature GLCM-variance from the DCE peak are in the top 10 for both training and test datasets. Additionally, the standard deviation of MRDI parameter  $A$  is in the top 10 for the SVM for both training and test datasets.



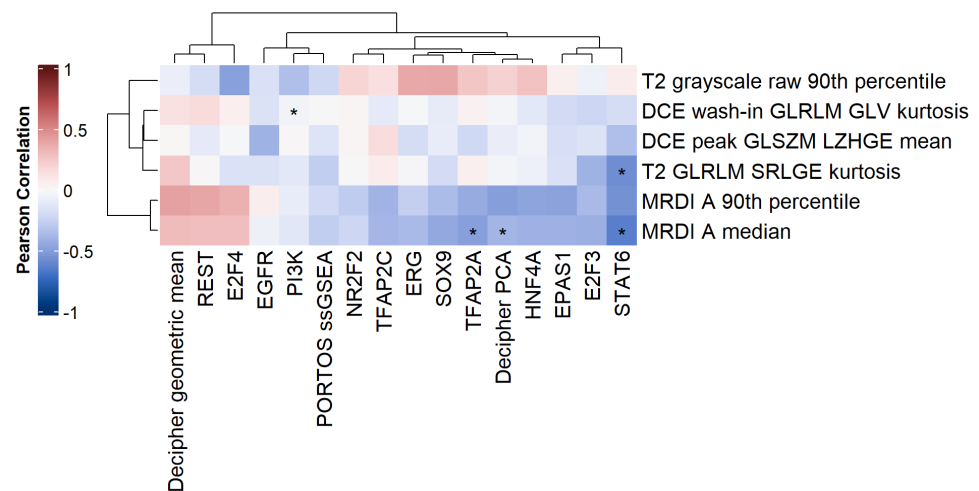
**Figure 5.** Normalized PFIs for LR, SVM, and kNN imaging classifiers on the training dataset. Features shared by all classifiers are shown in orange, the ones shared between LR and SVM are shown in blue, and the ones shared between SVM and kNN are shown in green. Features that are not shared are displayed in gray. The bars represent the standard deviations for the 10 random permutations.



**Figure 6.** Normalized PFIs for LR, SVM, and kNN imaging classifiers on the test dataset. Features shared by all classifiers are shown in orange, the ones shared between LR and SVM are shown in blue, the ones shared between SVM and kNN are shown in green, and the ones shared by LR and kNN are shown in purple. Features that are not shared are displayed in gray. The bars represent the standard deviations for the 10 random permutations.

### 3.5. Correlation with Transcriptomic Features

The heatmap of the Pearson correlation coefficient is shown in Figure 7, with significant correlations (adjusted  $p < 0.05$ ) highlighted by an asterisk. There were 5 significant correlations, of which 3 had moderate strength ( $\geq 0.5$ ). The strongest significant correlations were seen between the MRDI A median and the activity of the transcription factors STAT6 ( $-0.64$ ) and TFAP2A ( $-0.50$ ). The T2W GLRLM SRLGE kurtosis textural feature was also significantly correlated to the transcription factor STAT6 ( $-0.58$ ).



**Figure 7.** Heatmap of the Pearson correlation coefficient between imaging and transcriptomic features. Only coefficients with  $p$ -value  $\leq 0.2$  are shown. \*  $p < 0.05$ .

## 4. Discussion

The literature on PCa radiogenomics is scarce [21], and to our knowledge, none has explored the relationship between transcriptomics and quantitative radiomic features extracted from all three mpMRI sequences derived from a multicenter dataset. The identification of radiomic features describing aggressive PCa phenotypes can pave the way for providing non-invasive advice on patient-tailored treatment strategies. Our results show that spatiotemporal and textural features describing PCa anatomy and perfusion characteristics can be (cor)related to the underlying transcriptomic landscape.

The imaging machine's learning framework employed a moving window approach to extract texture features from T2W, ADC, and relevant time points of the dynamic DCE series. This method defines the ROI as a moving window, enabling the extraction of texture feature values at each imaging pixel, and resulting in a complete parametric map. In this way, it is possible to preserve the spatial characteristics of the features and retain potential texture heterogeneity within the image. In fact, our PFI analysis demonstrated that statistics such as kurtosis and standard deviation, which are related to the spatial distribution of the feature values, are more important than the means and median in the classification of clinically significant cancer, providing support for the adopted strategy. The best radiomic feature classifier was LR, which obtained a fair performance (AUC = 0.74) in predicting clinically significant PCa on an independent dataset. All models resulted in a high specificity ( $>0.80$ ), but poor sensitivity. From a clinical point of view, this is undesirable, as clinically significant PCa might be missed. This is likely related to the naturally occurring data imbalance in our datasets, which is a result of center-specific practices (i.e., which patients are treated with prostatectomy). Some of the main limitations of this study include the relatively small datasets used for the training of the imaging models, and more importantly, in the fact that they are imbalanced, such that the ratio of patients per group is inverse in the train and test datasets. In the imaging train dataset, 33% of the ROIs are of clinically significant PCa, while for the test dataset, this value is 69%. To cope with the data imbalance, for both LR and SVM classifiers, the class weight was set to "balanced", which automatically adjusts

weights to be inversely proportional to the class frequencies. Such an approach was not possible with KNN, which might explain this classifier's poor performance. The small sample size is also a limitation considering the optimal number of features to be used and might be responsible for some of the variability observed in terms of performance, as we are currently at the boundary of what can be explored with a dataset of this size. Larger datasets would allow for the validation of the results here presented as well as an investigation of a larger number of features. Additionally, hyperparameter optimization was performed using AUC as an optimization metric, as it leverages both sensitivity and specificity and deals in this way with the imbalance in the datasets. In addition to coping with the imbalanced dataset, this choice is also motivated by the clinical desire to optimize both sensitivity and specificity. At last, we performed a k-fold cross-validation scheme within the training dataset, repeating the procedure for 20 different random splits of the tumors in the 5 folds, so as to reduce the dependence of the estimated performance on the individual random split. Our AUC performance is slightly lower than reported in studies using the same Gleason score-based stratification (AUC~0.80) [21], but it is important to stress that ours is a multicenter study, likely incorporating larger variabilities in the measurement data. Radiomic features are known to be sensitive to image characteristics [63]; thus, by incorporating data from multiple institutes, scanners, and field strengths, we aimed to identify features that are robust to image variations.

The model-agnostic feature importance analysis using PFI showed that the highest-ranking imaging features overlapped between classifiers. Looking at the top four ranking features, three of them were shared between LR and SVM, and two between LR and kNN. Three features were shared within the top five for all classifiers: MRDI  $A$  std, T2 GLSZM LZLGE kurtosis, and the DCE wash-in NGTDM strength std, highlighting the importance of texture and spatiotemporal perfusion-based features for the stratification of clinically significant PCa (Figure 5). Despite the differences between our test and train datasets, 4 out of the top 10 highest-ranking radiomic features were common in both the train and test datasets, with T2 GLSZM LZLGE kurtosis appearing again as one of the overlapping features (Figure 6). The GLSZM features are rotation-independent and useful in evaluating non-periodic and heterogeneous textures as they measure consecutive groups of pixels with the same intensity, regardless of the direction. The large zone low gray-level emphasis (LZLGE) feature computes the zone distribution of the upper right quadrant of the GLSZM, with higher values of LZLGE being correlated to a greater number of large zones with low intensities [22]. The kurtosis is the fourth statistical moment of a probability distribution and can be interpreted as a measure of the relative weight of the tails, taking the normal distribution as a reference. Large kurtosis indicates that tail values are more extreme compared to a normal distribution [22]. In our context, a small number of very low or very high values of large groups of pixels with lower intensities might be observed in the investigated tumors. Low T2W signal intensity is a characteristic trait of PCa, and very large regions of low gray-level values could represent larger tumors and more aggressive PCa. MRDI  $A$  is given by the product of  $K^{trans}$  and  $\alpha$ , with  $\alpha$  being equal to the time integral of the intravascular concentration, and reflecting both extravasation and perfusion [20,39], which are typically increased due to angiogenesis. A higher standard deviation on this parameter's values represents a more heterogeneous pattern of enhancement within the tumor. Interestingly, while ADC values have often been shown to inversely correlate with ISUP grade [64,65], in our study, there were no ADC features present in the top 10 for any of the 3 imaging classifiers. In line with our findings, when using a multicenter MRI dataset, Bengtsson et al. [66] were not able to identify a correlation between ADC and ISUP grades, suggesting that the lack of harmonization in MRI parameters used to acquire and derive ADC maps might dilute the value of this feature in a clinical situation involving different field strengths, scanners, and acquisition parameters. This could potentially be an explanation for our results, which made use of a multicenter training dataset to identify the most important features. The highest-ranking imaging features suggest that T2W, DCE wash-ins, peak frames, as well as MRDI maps, contain relevant features for

risk stratification. For the goal of the present study, the highest-ranking imaging features using the LR model were used to further investigate the relationship between imaging and transcriptomic data.

Our results established a significant relationship between imaging features and transcriptomic characteristics. A set of 31 imaging features was correlated with 33 transcriptomic features, yielding 5 significant correlations, 3 of which were of moderate strength ( $|r| \geq 0.5$ )—Figure 7. The strongest significant correlations were seen between the perfusion-based imaging feature—MRDI *A* median—and the activity of the transcription factors STAT6 (−0.64) and TFAP2A (−0.50). A higher-order T2W textural feature was significantly correlated to the transcription factor STAT6 (−0.58). The STAT6 expression plays a role in PCa cell proliferation and migration and was found to be positively correlated with the tumor size and high PCa histological grades [67]. TFAP2A is a well-characterized transcription factor that regulates proliferation and differentiation in mammalian cells [68]. Conversely, the loss of the AP2alpha protein expression has been strongly associated with the aggressiveness and progression of numerous types of cancers, including prostate cancer [69]. In addition, we observed a significant correlation between the MRDI *A* median and the gene signature Decipher PCA (−0.40), as well as between the pathway PI3K and DCE wash-in GLRLM GLV kurtosis (−0.04), which were below the set threshold of  $|r| \geq 0.5$ . In the study by Hectors et al. [35], a moderate positive correlation ( $r = 0.40$ ) was previously reported between a T2w GLCM texture feature and the Decipher risk score. It is important to note that the way the Decipher score is derived in the current study differs from the Decipher risk scores typically reported in the literature, derived using clinical-grade whole-transcriptome Decipher assays. Here, the Decipher score exhibits a moderate correlation with the parameter MRDI *A*, derived from DCE-MRI. Microvessel density has been associated with recurrence and prognosis in patients with PCa [70], and angiogenesis is an essential step for tumor growth and the initiation of metastasis [71]. Oto et al. [72] observed a significant moderate correlation ( $r = 0.44$ ) between microvessel density and the DCE-MRI parameter  $k_{ep}$  in prostate cancer (which describes the back flux rate constant between the extracellular space and the plasma). Previous findings, thus, provide support to our result that MRDI *A*, which quantitatively describes perfusion and extravasation associated with angiogenesis, is related to the Decipher score, which grades tumor aggressiveness. The PI3K pathway is involved in cellular proliferation and growth, and an elevated PI3K signaling is considered a hallmark of cancer [73]. Thus, it would not be surprising to find a correlation between PI3K and DCE-MRI features, as this MRI sequence is used for oncological purposes, precisely to highlight angiogenesis processes. In our study, despite the significance, the correlation we observe between PI3K and the DCE wash-in GLRLM GLV kurtosis is close to zero in value. The strength of the significant correlations observed between imaging and genomic features is higher in our study when compared to the values described in a similar study by Hectors et al. [35]. The analysis of immune cell-type fractions was not addressed in this manuscript as they are not easily interpretable in this context. Single-cell RNA-seq is a strong unbiased method used to identify specific tumor cell subpopulations. It can reveal cellular heterogeneity and possibly identify distinct cell populations in the tumor microenvironment [74], and it is an interesting alternative to bulk RNA-seq data that would likely boost the performance of transcriptomic feature-based models. However, MRI is, by default, a bulk low-resolution method, and in this regard, bulk RNA-seq provides more suitable data to match it. Alternatively, spatial transcriptomics [75] retrieves spatially resolved maps of gene expression profiles and can investigate tumor heterogeneity, which can be (cor)related to the MR imaging domain.

The aim of this work was to identify candidate imaging features that can potentially represent relevant phenotypes for individualized patient treatment. Our strategy consisted of a machine learning framework to stratify patients based on their ISUP grade. Although relevant and typically investigated, the ISUP grade is not a complete depiction of the complexity of PCa, and a more suitable clinical outcome, such as 5-year biochemical recurrence-free or metastasis-free survival, would have been preferred. Nonetheless,

datasets encompassing this level of detailed clinical information, which require long follow-up times in combination with MRIs (and genomic data), are virtually impossible to obtain, retrospectively. In addition, the use of prostatectomy specimens as the reference standard facilitated a more accurate localization of pathology-proven lesions on the MRI grid, but it also limited the sample size and potentially biased our study towards the subgroup of surgically treated PCa patients. Unlike other imaging modalities, MRI intensities have arbitrary units. This fact alone can significantly affect the absolute values of the derived radiomic features and, thus, their robustness. The robustness of radiomic features in a multi-center setting is a prerequisite for the successful translation of radiomic biomarkers in a clinical workflow [76]. In order to mitigate these issues, intensity normalization of non-quantitative images (T2W and DCE) was performed based on the prostate region prior to feature extraction. Alternative methods for image normalization and their impacts on radiomic feature robustness have been investigated [77]; histogram matching seems to be the most promising normalization scheme. A future prospective study would enable data harmonization and the validation of our findings in a larger cohort.

## 5. Conclusions

In this study, we identified radiomic features derived from mpMRI that are predictive of clinically significant PCa and are significantly correlated with some of the corresponding transcriptomic features. Our results suggest that mpMRI imaging features hold the potential to decode tumor phenotypes, paving the way for a non-invasive assessment of PCa and more individualized treatment regimens.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15123074/s1>, Table S1: Overview of MRI acquisition settings in the training dataset; Table S2: Overview of MRI acquisition settings in the test dataset; Table S3: Complete list of transcriptomic features consisting of 132 functional features describing 14 intracellular signaling pathways and 118 transcription factor activities, as well as 6 gene signature scores.

**Author Contributions:** Conceptualization, S.T., C.D.F., H.W., F.E., P.v.H., U.v.d.H. and M.M.; methodology, S.T., F.E., C.D.F., H.W., W.Z., U.v.d.H., P.v.H. and M.M.; software, A.S., J.K., S.T. and C.D.F.; validation, S.T. and C.D.F.; formal analysis, A.S. and J.K.; investigation, A.S., J.K., S.T., C.D.F. and F.E.; resources, P.v.H., W.Z., S.L., E.B., A.M.B. and U.v.d.H.; data curation, A.S. and J.K.; writing—original draft preparation, A.S., J.K., C.D.F. and S.T.; writing—review and editing, H.W., F.E., W.Z., U.v.d.H., P.v.H., S.L., E.B., A.M.B. and M.M.; visualization, A.S. and J.K.; supervision, S.T., F.E., H.W., C.D.F. and M.M.; project administration, S.T.; funding acquisition, S.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Hanarth Fonds fellowship for AI in Oncology, granted to S.T. in 2019.

**Institutional Review Board Statement:** All data included in this study were obtained from studies that were conducted in accordance with the Declaration of Helsinki. The Prostate Cancer Molecular Medicine study was approved by the Institutional Review Board of the Erasmus Medical Center, Rotterdam, The Netherlands (METC 2010-176, NL32105.078.10, 1st of June 2010). The patient data with MRI and RNA sequencing information were collected after approval by the Ethics Committee of The Netherlands Cancer Institute.

**Informed Consent Statement:** Written informed consent was obtained from all patients to use their data for research.

**Data Availability Statement:** Data are not publicly available due to GDPR protection and data-sharing agreements.

**Acknowledgments:** The authors would like to thank Jelle Barenstz and Stijn Heijmink for providing the data.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **2021**, *149*, 778–789. [CrossRef] [PubMed]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
3. EAU Guidelines on Prostate Cancer. Available online: <https://uroweb.org/guidelines/prostate-cancer> (accessed on 4 April 2023).
4. Mottet, N.; van den Bergh, R.C.; Briers, E.; Van den Broeck, T.; Cumberbatch, M.G.; De Santis, M.; Fanti, S.; Fossati, N.; Gandaglia, G.; Gillessen, S.; et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer—2020 update. Part 1: Screening, diagnosis, and local treatment with curative intent. *Eur. Urol.* **2021**, *79*, 243–262. [CrossRef] [PubMed]
5. Stabile, A.; Giganti, F.; Rosenkrantz, A.B.; Taneja, S.S.; Villeirs, G.; Gill, I.S.; Allen, C.; Emberton, M.; Moore, C.M.; Kasivisvanathan, V. Multiparametric MRI for prostate cancer diagnosis: Current status and future directions. *Nat. Rev. Urol.* **2020**, *17*, 41–61. [CrossRef]
6. Van Poppel, H.; Roobol, M.J.; Chapple, C.R.; Catto, J.W.; N'Dow, J.; Sønksen, J.; Stenzl, A.; Wirth, M. Prostate-specific antigen testing as part of a risk-adapted early detection strategy for prostate cancer: European Association of Urology position and recommendations for 2021. *Eur. Urol.* **2021**, *80*, 703–711. [CrossRef] [PubMed]
7. Jairath, N.K.; Dal Pra, A.; Vince Jr, R.; Dess, R.T.; Jackson, W.C.; Tosoian, J.J.; McBride, S.M.; Zhao, S.G.; Berlin, A.; Mahal, B.A.; et al. A systematic review of the evidence for the decipher genomic classifier in prostate cancer. *Eur. Urol.* **2021**, *79*, 374–383. [CrossRef]
8. Fine, N.D.; LaPolla, F.; Epstein, M.; Loeb, S.; Dani, H. Genomic classifiers for treatment selection in newly diagnosed prostate cancer. *BJU Int.* **2019**, *124*, 578–586. [CrossRef]
9. Eastham, J.A.; Auffenberg, G.B.; Barocas, D.A.; Chou, R.; Crispino, T.; Davis, J.W.; Eggener, S.; Horwitz, E.M.; Kane, C.J.; Kirkby, E.; et al. Clinically localized prostate cancer: AUA/ASTRO guideline, part I: Introduction, risk assessment, staging, and risk-based management. *J. Urol.* **2022**, *208*, 10–18. [CrossRef]
10. Merseburger, A.S.; Waldron, N.; Ribal, M.J.; Heidenreich, A.; Perner, S.; Fizazi, K.; Sternberg, C.N.; Mateo, J.; Wirth, M.P.; Castro, E.; et al. Genomic testing in patients with metastatic castration-resistant prostate cancer: A pragmatic guide for clinicians. *Eur. Urol.* **2021**, *79*, 519–529. [CrossRef]
11. Xiang, J.; Yan, H.; Li, J.; Wang, X.; Chen, H.; Zheng, X. Transperineal versus transrectal prostate biopsy in the diagnosis of prostate cancer: A systematic review and meta-analysis. *World J. Surg. Oncol.* **2019**, *17*, 1–11. [CrossRef]
12. Sun, Y.; Reynolds, H.M.; Parameswaran, B.; Wraith, D.; Finnegan, M.E.; Williams, S.; Haworth, A. Multiparametric MRI and radiomics in prostate cancer: A review. *Australas. Phys. Eng. Sci. Med.* **2019**, *42*, 3–25. [CrossRef] [PubMed]
13. Drost, F.J.H.; Osses, D.; Nieboer, D.; Bangma, C.H.; Steyerberg, E.W.; Roobol, M.J.; Schoots, I.G. Prostate magnetic resonance imaging, with or without magnetic resonance imaging-targeted biopsy, and systematic biopsy for detecting prostate cancer: A Cochrane systematic review and meta-analysis. *Eur. Urol.* **2020**, *77*, 78–94. [CrossRef] [PubMed]
14. Eastham, J.A.; Auffenberg, G.B.; Barocas, D.A.; Chou, R.; Crispino, T.; Davis, J.W.; Eggener, S.; Horwitz, E.M.; Kane, C.J.; Kirkby, E.; et al. Clinically localized prostate cancer: AUA/ASTRO guideline, part II: principles of active surveillance, principles of surgery, and follow-up. *J. Urol.* **2022**, *208*, 19–25. [CrossRef] [PubMed]
15. Barrett, T.; Rajesh, A.; Rosenkrantz, A.B.; Choyke, P.L.; Turkbey, B. PI-RADS version 2.1: One small step for prostate MRI. *Clin. Radiol.* **2019**, *74*, 841–852. [CrossRef] [PubMed]
16. Wang, J.; Wu, C.J.; Bao, M.L.; Zhang, J.; Wang, X.N.; Zhang, Y.D. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur. Radiol.* **2017**, *27*, 4082–4090. [CrossRef]
17. Hambrock, T.; Vos, P.C.; Hulsbergen-van de Kaa, C.A.; Barentsz, J.O.; Huisman, H.J. Prostate cancer: Computer-aided diagnosis with multiparametric 3-T MR imaging—Effect on observer performance. *Radiology* **2013**, *266*, 521–530. [CrossRef] [PubMed]
18. Meyer, H.J.; Wienke, A.; Surov, A. Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient—A systematic review and meta analysis. *BMC Cancer* **2020**, *20*, 482. [CrossRef]
19. Franiel, T.; Hamm, B.; Hricak, H. Dynamic contrast-enhanced magnetic resonance imaging and pharmacokinetic models in prostate cancer. *Eur. Radiol.* **2011**, *21*, 616–626. [CrossRef]
20. Turco, S.; Lavini, C.; Heijmink, S.; Barentsz, J.; Wijkstra, H.; Mischi, M. Evaluation of dispersion MRI for improved prostate cancer diagnosis in a multicenter study. *AJR Am. J. Roentgenol.* **2018**, *211*, 242–251. [CrossRef]
21. Ferro, M.; de Cobelli, O.; Vartolomei, M.D.; Lucarelli, G.; Crocetto, F.; Barone, B.; Sciarra, A.; Del Giudice, F.; Muto, M.; Maggi, M.; et al. Prostate cancer radiogenomics—From imaging to molecular characterization. *Int. J. Mol. Sci.* **2021**, *22*, 9971. [CrossRef] [PubMed]
22. Ardakani, A.A.; Bureau, N.J.; Ciaccio, E.J.; Acharya, U.R. Interpretation of radiomics features—A pictorial review. *Comput. Methods Programs Biomed.* **2022**, *215*, 106609. [CrossRef] [PubMed]

23. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)] [[PubMed](#)]
24. Lambin, P.; Leijenaar, R.T.; Deist, T.M.; Peerlings, J.; De Jong, E.E.; Van Timmeren, J.; Sanduleanu, S.; Larue, R.T.; Even, A.J.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)]
25. Bodalal, Z.; Trebeschi, S.; Nguyen-Kim, T.D.L.; Schats, W.; Beets-Tan, R. Radiogenomics: Bridging imaging and genomics. *Abdom. Radiol.* **2019**, *44*, 1960–1984. [[CrossRef](#)]
26. Perlis, N.; Al-Kasab, T.; Ahmad, A.; Goldberg, E.; Fadak, K.; Sayid, R.; Finelli, A.; Kulkarni, G.; Hamilton, R.; Zlotta, A.; et al. Defining a cohort that may not require repeat prostate biopsy based on PCA3 score and magnetic resonance imaging: The dual negative effect. *J. Urol.* **2018**, *199*, 1182–1187. [[CrossRef](#)] [[PubMed](#)]
27. Fenstermaker, M.; Mendhiratta, N.; Bjurlin, M.A.; Meng, X.; Rosenkrantz, A.B.; Huang, R.; Deng, F.M.; Zhou, M.; Huang, W.C.; Lepor, H.; et al. Risk stratification by urinary prostate cancer gene 3 testing before magnetic resonance imaging-ultrasound fusion-targeted prostate biopsy among men with no history of biopsy. *Urology* **2017**, *99*, 174–179. [[CrossRef](#)]
28. De Luca, S.; Passera, R.; Cattaneo, G.; Manfredi, M.; Mele, F.; Fiori, C.; Bollito, E.; Cirillo, S.; Porpiglia, F. High prostate cancer gene 3 (PCA 3) scores are associated with elevated Prostate Imaging Reporting and Data System (PI-RADS) grade and biopsy Gleason score, at magnetic resonance imaging/ultrasonography fusion software-based targeted prostate biopsy after a previous negative standard biopsy. *BJU Int.* **2016**, *118*, 723–730. [[PubMed](#)]
29. Martin, D.T.; Ghabili, K.; Levi, A.; Humphrey, P.A.; Sprenkle, P.C. Prostate cancer genomic classifier relates more strongly to Gleason grade group than Prostate Imaging Reporting and Data System score in multiparametric prostate magnetic resonance imaging-ultrasound fusion targeted biopsies. *Urology* **2019**, *125*, 64–72. [[CrossRef](#)]
30. Leapman, M.S.; Westphalen, A.C.; Ameli, N.; Lawrence, H.J.; Febbo, P.G.; Cooperberg, M.R.; Carroll, P.R. Association between a 17-gene genomic prostate score and multi-parametric prostate MRI in men with low and intermediate risk prostate cancer (PCa). *PLoS ONE* **2017**, *12*, e0185535. [[CrossRef](#)] [[PubMed](#)]
31. Beksac, A.T.; Cumarasamy, S.; Falagarino, U.; Xu, P.; Takhar, M.; Alshalalfa, M.; Gupta, A.; Prasad, S.; Martini, A.; Thulasidass, H.; et al. Multiparametric magnetic resonance imaging features identify aggressive prostate cancer at the phenotypic and transcriptomic level. *J. Urol.* **2018**, *200*, 1241–1249. [[CrossRef](#)] [[PubMed](#)]
32. Salmasi, A.; Said, J.; Shindel, A.W.; Khoshnoodi, P.; Felker, E.R.; Sisk, A.E., Jr.; Grogan, T.; McCullough, D.; Bennett, J.; Bailey, H.; et al. A 17-gene genomic prostate score assay provides independent information on adverse pathology in the setting of combined multiparametric magnetic resonance imaging fusion targeted and systematic prostate biopsy. *J. Urol.* **2018**, *200*, 564–572. [[CrossRef](#)] [[PubMed](#)]
33. McCann, S.M.; Jiang, Y.; Fan, X.; Wang, J.; Antic, T.; Prior, F.; VanderWeele, D.; Oto, A. Quantitative multiparametric MRI features and PTEN expression of peripheral zone prostate cancer: A pilot study. *Am. J. Roentgenol.* **2016**, *206*, 559–565. [[CrossRef](#)]
34. Stoyanova, R.; Pollack, A.; Takhar, M.; Lynne, C.; Parra, N.; Lam, L.L.; Alshalalfa, M.; Buerki, C.; Castillo, R.; Jorda, M.; et al. Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies. *Oncotarget* **2016**, *7*, 53362. [[CrossRef](#)]
35. Hectors, S.J.; Cherny, M.; Yadav, K.K.; Beksaç, A.T.; Thulasidass, H.; Lewis, S.; Davicioni, E.; Wang, P.; Tewari, A.K.; Taouli, B. Radiomics features measured with multiparametric magnetic resonance imaging predict prostate cancer aggressiveness. *J. Urol.* **2019**, *202*, 498–505. [[CrossRef](#)]
36. Switlyk, M.D.; Salberg, U.B.; Geier, O.M.; Vlatkovic, L.; Lilleby, W.; Lyng, H.; Seierstad, T. PTEN expression in prostate cancer: Relationship with clinicopathologic features and multiparametric MRI findings. *Am. J. Roentgenol.* **2019**, *212*, 1206–1214. [[CrossRef](#)] [[PubMed](#)]
37. Fischer, S.; Tahoun, M.; Kwaan, B.; Thierfelder, K.M.; Weber, M.A.; Krause, B.J.; Hakenberg, O.; Fuellen, G.; Hamed, M. A radiogenomic approach for decoding molecular mechanisms underlying tumor progression in prostate cancer. *Cancers* **2019**, *11*, 1293. [[CrossRef](#)] [[PubMed](#)]
38. Sun, Y.; Williams, S.; Byrne, D.; Keam, S.; Reynolds, H.M.; Mitchell, C.; Wraith, D.; Murphy, D.; Haworth, A. Association analysis between quantitative MRI features and hypoxia-related genetic profiles in prostate cancer: A pilot study. *Br. J. Radiol.* **2019**, *92*, 20190373. [[CrossRef](#)] [[PubMed](#)]
39. Mischi, M.; Turco, S.; Lavini, C.; Kompatsiari, K.; de la Rosette, J.J.; Breeuwer, M.; Wijkstra, H. Magnetic resonance dispersion imaging for localization of angiogenesis and cancer growth. *Investig. Radiol.* **2014**, *49*, 561–569. [[CrossRef](#)] [[PubMed](#)]
40. Abeshouse, A.; Ahn, J.; Akbani, R.; Ally, A.; Amin, S.; Andry, C.D.; Annala, M.; Aprikian, A.; Armenia, J.; Arora, A.; et al. The molecular taxonomy of primary prostate cancer. *Cell* **2015**, *163*, 1011–1025. [[CrossRef](#)] [[PubMed](#)]
41. Srigley, J.R.; Delahunt, B.; Samaratunga, H.; Billis, A.; Cheng, L.; Clouston, D.; Evans, A.; Furusato, B.; Kench, J.; Leite, K.; et al. Controversial issues in Gleason and International Society of Urological Pathology (ISUP) prostate cancer grading: Proposed recommendations for international implementation. *Pathology* **2019**, *51*, 463–473. [[CrossRef](#)]
42. Weinreb, J.C.; Barentsz, J.O.; Choyke, P.L.; Cornud, F.; Haider, M.A.; Macura, K.J.; Margolis, D.; Schnall, M.D.; Shtern, F.; Tempany, C.M.; et al. PI-RADS prostate imaging—Reporting and data system: 2015, version 2. *Eur. Urol.* **2016**, *69*, 16–40. [[CrossRef](#)] [[PubMed](#)]

43. Dinis Fernandes, C.; Turco, S.; Mischi, M.; Wijkstra, H. Deep learning for prostate and zonal segmentation on a multicenter MRI dataset. In Proceedings of the 9th EAU Section of Urological Imaging 2021, Athens, Greece, 25 November 2021.
44. Yushkevich, P.A.; Piven, J.; Cody Hazlett, H.; Gimpel Smith, R.; Ho, S.; Gee, J.C.; Gerig, G. User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage* **2006**, *31*, 1116–1128. [[CrossRef](#)] [[PubMed](#)]
45. Finotello, F.; Mayer, C.; Plattner, C.; Laschober, G.; Rieder, D.; Hackl, H.; Krogsdam, A.; Loncova, Z.; Posch, W.; Wilflingseder, D.; Sopper, S.; et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **2019**, *11*, 1–20.
46. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
47. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
48. Lapuente-Santana, Ó.; van Genderen, M.; Hilbers, P.A.; Finotello, F.; Eduati, F. Interpretable systems biomarkers predict response to immune-checkpoint inhibitors. *Patterns* **2021**, *2*, 100293. [[CrossRef](#)] [[PubMed](#)]
49. Schubert, M.; Klinger, B.; Klünemann, M.; Sieber, A.; Uhlitz, F.; Sauer, S.; Garnett, M.J.; Blüthgen, N.; Saez-Rodriguez, J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **2018**, *9*, 20. [[CrossRef](#)]
50. Garcia-Alonso, L.; Holland, C.H.; Ibrahim, M.M.; Turei, D.; Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **2019**, *29*, 1363–1375. [[CrossRef](#)]
51. Zhao, S.G.; Chang, S.L.; Spratt, D.E.; Erho, N.; Yu, M.; Ashab, H.A.D.; Alshalalfa, M.; Speers, C.; Tomlins, S.A.; Davicioni, E.; et al. Development and validation of a 24-gene predictor of response to postoperative radiotherapy in prostate cancer: A matched, retrospective analysis. *Lancet Oncol.* **2016**, *17*, 1612–1620. [[CrossRef](#)] [[PubMed](#)]
52. Erho, N.; Crisan, A.; Vergara, I.A.; Mitra, A.P.; Ghadessi, M.; Buerki, C.; Bergstralh, E.J.; Kollmeyer, T.; Fink, S.; Haddad, Z.; et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS ONE* **2013**, *8*, e66855. [[CrossRef](#)] [[PubMed](#)]
53. Hänzelmann, S.; Castelo, R.; Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **2013**, *14*, 7. [[CrossRef](#)] [[PubMed](#)]
54. Rooney, M.S.; Shukla, S.A.; Wu, C.J.; Getz, G.; Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **2015**, *160*, 48–61. [[CrossRef](#)]
55. Messina, J.L.; Fenstermacher, D.A.; Eschrich, S.; Qu, X.; Berglund, A.E.; Lloyd, M.C.; Schell, M.J.; Sondak, V.K.; Weber, J.S.; Mulé, J.J. 12-Chemokine gene signature identifies lymph node-like structures in melanoma: Potential for patient selection for immunotherapy? *Sci. Rep.* **2012**, *2*, 765. [[CrossRef](#)] [[PubMed](#)]
56. Szumilas, M. Explaining odds ratios. *J. Can. Acad. Child Adolesc. Psychiatry* **2010**, *19*, 227. [[PubMed](#)]
57. Zhao, Z.; Anand, R.; Wang, M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Washington, DC, USA, 5–8 October 2019; pp. 442–452.
58. Verma, S.; Turkbey, B.; Muradyan, N.; Rajesh, A.; Cornud, F.; Haider, M.A.; Choyke, P.L.; Harisinghani, M. Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management. *AJR. Am. J. Roentgenol.* **2012**, *198*, 1277. [[CrossRef](#)]
59. Vallieres, M.; Kay-Rivest, E.; Perrin, L.J.; Liem, X.; Furstoss, C.; Aerts, H.J.; Khaouam, N.; Nguyen-Tan, P.F.; Wang, C.S.; Sultanem, K.; et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci. Rep.* **2017**, *7*, 10117. [[CrossRef](#)] [[PubMed](#)]
60. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
61. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
62. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; Volume 57, pp. 10–25080.
63. Roy, S.; Whitehead, T.D.; Quirk, J.D.; Salter, A.; Ademuyiwa, F.O.; Li, S.; An, H.; Shoghi, K.I. Optimal co-clinical radiomics: Sensitivity of radiomic features to tumour volume, image noise and resolution in co-clinical T1-weighted and T2-weighted magnetic resonance imaging. *EBioMedicine* **2020**, *59*, 102963. [[CrossRef](#)] [[PubMed](#)]
64. Boesen, L.; Chabanova, E.; Løgager, V.; Balslev, I.; Thomsen, H.S. Apparent diffusion coefficient ratio correlates significantly with prostate cancer gleason score at final pathology. *J. Magn. Reson. Imaging JMRI* **2015**, *42*, 446–453. [[CrossRef](#)] [[PubMed](#)]
65. Verma, S.; Rajesh, A.; Morales, H.; Lemen, L.; Bills, G.; Delworth, M.; Gaitonde, K.; Ying, J.; Samartunga, R.; Lamba, M. Assessment of aggressiveness of prostate cancer: Correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR. Am. J. Roentgenol.* **2011**, *196*, 374–381. [[CrossRef](#)] [[PubMed](#)]
66. Bengtsson, J.; Thimansson, E.; Baubeta, E.; Zackrisson, S.; Sundgren, P.C.; Bjartell, A.; Flondell-Sité, D. Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners. *Front. Oncol.* **2023**, *13*, 1079040. [[CrossRef](#)] [[PubMed](#)]
67. Das, S.; Roth, C.P.; Wasson, L.M.; Vishwanatha, J.K. Signal transducer and activator of transcription-6 (STAT6) is a constitutively expressed survival factor in human prostate cancer. *Prostate* **2007**, *67*, 1550–1564. [[CrossRef](#)] [[PubMed](#)]

68. Schorle, H.; Meier, P.; Buchert, M.; Jaenisch, R.; Mitchell, P.J. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **1996**, *381*, 235–238. [[CrossRef](#)]
69. Lipponen, P.; Aaltomaa, S.; Kellokoski, J.; Ala-Opas, M.; Kosma, V.M. Expression of activator protein 2 in prostate cancer is related to tumor differentiation and cell proliferation. *Eur. Urol.* **2000**, *37*, 573–578. [[CrossRef](#)] [[PubMed](#)]
70. de la Taille, A.; Katz, A.E.; Bagiella, E.; Buttyan, R.; Sharir, S.; Olsson, C.A.; Burchardt, T.; Ennis, R.D.; Rubin, M.A. Microvessel density as a predictor of PSA recurrence after radical prostatectomy: A comparison of CD34 and CD31. *Am. J. Clin. Pathol.* **2000**, *113*, 555–562. [[CrossRef](#)]
71. Miller, J.C.; Pien, H.H.; Sahani, D.; Sorensen, A.G.; Thrall, J.H. Imaging angiogenesis: Applications and potential for drug development. *J. Natl. Cancer Inst.* **2005**, *97*, 172–187. [[CrossRef](#)] [[PubMed](#)]
72. Oto, A.; Yang, C.; Kayhan, A.; Tretiakova, M.; Antic, T.; Schmid-Tannwald, C.; Eggener, S.; Karczmar, G.S.; Stadler, W.M. Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: Correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. *Am. J. Roentgenol.* **2011**, *197*, 1382–1390. [[CrossRef](#)] [[PubMed](#)]
73. Fruman, D.A.; Chiu, H.; Hopkins, B.D.; Bagrodia, S.; Cantley, L.C.; Abraham, R.T. The PI3K pathway in human disease. *Cell* **2017**, *170*, 605–635. [[CrossRef](#)]
74. Wong, H.Y.; Sheng, Q.; Hesterberg, A.B.; Croessmann, S.; Rios, B.L.; Giri, K.; Jackson, J.; Miranda, A.X.; Watkins, E.; Schaffer, K.R.; et al. Single cell analysis of cribriform prostate cancer reveals cell intrinsic and tumor microenvironmental pathways of aggressive disease. *Nat. Commun.* **2022**, *13*, 6036. [[CrossRef](#)]
75. Berglund, E.; Maaskola, J.; Schultz, N.; Friedrich, S.; Marklund, M.; Bergensträhle, J.; Tarish, F.; Tanoglidi, A.; Vickovic, S.; Larsson, L.; et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **2018**, *9*, 2419. [[CrossRef](#)] [[PubMed](#)]
76. Fournier, L.; Costaridou, L.; Bidaut, L.; Michoux, N.; Lecouvet, F.E.; de Geus-Oei, L.F.; Boellaard, R.; Oprea-Lager, D.E.; Obuchowski, N.A.; Caroli, A.; et al. Incorporating radiomics into clinical trials: Expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur. Radiol.* **2021**, *31*, 6001–6012. [[CrossRef](#)] [[PubMed](#)]
77. Saltybaeva, N.; Tanadini-Lang, S.; Vuong, D.; Burgermeister, S.; Mayinger, M.; Bink, A.; Andratschke, N.; Guckenberger, M.; Bogowicz, M. Robustness of radiomic features in magnetic resonance imaging for patients with glioblastoma: Multi-center study. *Phys. Imaging Radiat. Oncol.* **2022**, *22*, 131–136. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.