

Aus dem Institut für klinische Radiologie und Nuklearmedizin  
der Medizinischen Fakultät Mannheim  
Direktor: Prof. Dr. med. Schönberg

Entwicklung einer Radiomics-Signatur sowie eines Deep Learning Algorithmus  
für die Prädiktion von signifikanten Prostatakarzinomen

Inauguraldissertation  
zur Erlangung des medizinischen Doktorgrades  
der  
Medizinischen Fakultät Mannheim  
der Ruprecht-Karls-Universität  
zu  
Heidelberg

vorgelegt von  
Sarah Elena Schenk

aus  
Mannheim  
2022

Dekan: Prof. Dr. med. Sergij Goerd  
Referentin: Frau Univ.-Prof. Dr. med. Ulrike Attenberger

Für meine Familie.

# INHALTSVERZEICHNIS

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b><i>Abkürzungsverzeichnis</i></b>   | <b>5</b>  |
| <b>2</b>  | <b><i>Einleitung und Grundlagen</i></b>   | <b>7</b>  |
| 2.1.      | <b>Epidemiologie des Prostatakarzinoms</b>  | <b>7</b>  |
| 2.2.      | <b>Ätiologie und Risikofaktoren des Prostatakarzinoms</b>   | <b>7</b>  |
| 2.3.      | <b>Anatomie des Prostatakarzinoms</b>   | <b>8</b>  |
| 2.4.      | <b>Histopathologie des Prostatakarzinoms</b>  | <b>9</b>  |
| 2.4.1.    | Gleason-Score   | 10        |
| 2.5.      | <b>Klinik des Prostatakarzinoms</b>   | <b>11</b> |
| 2.6.      | <b>Klassifikation des Prostatakarzinoms</b>   | <b>11</b> |
| 2.7.      | <b>Diagnostik des Prostatakarzinoms</b>   | <b>12</b> |
| 2.7.1.    | Früherkennung: Prostata-spezifisches Antigen und digital-rektale Untersuchung   | 12        |
| 2.7.2.    | Transrektaler Ultraschall   | 13        |
| 2.7.3.    | Biopsie   | 13        |
| 2.7.4.    | Multiparametrische MRT der Prostata   | 14        |
| 2.7.5.    | Prostate Imaging – Reporting and Data System  | 18        |
| 2.8.      | <b>Klinische Problemstellung</b>  | <b>20</b> |
| 2.8.1.    | Undergrading des PCa  | 21        |
| 2.8.2.    | Überdiagnose des PCa  | 23        |
| 2.9.      | <b>Radiomics</b>  | <b>25</b> |
| 2.9.1.    | Radiomics: Begriff und Definition   | 25        |
| 2.9.2.    | Funktionsprinzip von Radiomics  | 26        |
| 2.9.3.    | Radiomics Feature-Klassen   | 27        |
| 2.10.     | <b>Künstliche Intelligenz: Grundprinzipien von Machine Learning</b>   | <b>28</b> |
| 2.10.1.   | Klassifikationsmodelle  | 29        |
| 2.11.     | <b>Methodologische Grundlagen zur Gewinnung von Radiomics Features und zur Entwicklung eines Machine Learning Modells</b> | <b>32</b> |
| 2.12.     | <b>Entwicklung eines Machine-Learning-Algorithmus: Radiomics Pipeline</b>   | <b>34</b> |
| 2.12.1.   | Datenakquisition  | 34        |
| 2.12.2.   | Segmentierung bzw. Region of Interest-Auswahl   | 35        |
| 2.12.3.   | Datenvorverarbeitung („Preprocessing“)  | 36        |
| 2.12.4.   | Datenaugmentierung und Balancing  | 37        |
| 2.12.5.   | Extraktion von Radiomics Features   | 38        |
| 2.12.6.   | Selektion von Radiomics Features  | 38        |
| 2.12.7.   | Training des Modells  | 39        |
| 2.12.8.   | Interne Validierung des Modells   | 39        |
| 2.12.9.   | Test und externe Validierung des Modells  | 40        |
| 2.12.10.  | Performance – Maße  | 40        |
| 2.13.     | <b>Problematik der Reproduzierbarkeit und Generalisierbarkeit von MRT-basierten Features</b>                              | <b>42</b> |
| 2.14.     | <b>Künstliches neuronales Netzwerk, Deep Learning und Faltungsnetz</b>  | <b>43</b> |
| 2.15.     | <b>Zielsetzung dieser Arbeit</b>  | <b>45</b> |
| <b>3.</b> | <b><i>Material und Methoden</i></b>   | <b>46</b> |
| 3.1.      | <b>Ethikvotum</b>   | <b>46</b> |
| 3.2.      | <b>Patientenkollektiv</b>   | <b>46</b> |

|       |  |           |
|-------|--|-----------|
| 3.3.  | <b>Referenzstandard: Histopathologische Korrelation</b>                                  | <b>48</b> |
| 3.4.  | <b>Segmentierung</b>   | <b>48</b> |
| 3.5.  | <b>Preprocessing</b>   | <b>50</b> |
| 3.6.  | <b>Datenaugmentierung</b>  | <b>51</b> |
| 3.7.  | <b>Radiomics Features - Extraktion und Selektion</b>                                     | <b>51</b> |
| 3.8.  | <b>Extrahierte Radiomics Features</b>  | <b>51</b> |
| 3.9.  | <b>Training-Validierungs-Split</b>   | <b>53</b> |
| 3.10. | <b>Klassifikationsmodelle: Selektion und Validierung</b>                                 | <b>53</b> |
| 3.11. | <b>Performance-Maße</b>  | <b>53</b> |
| 3.12. | <b>Alternativer Ansatz: Deep Learning</b>  | <b>53</b> |
| 3.13. | <b>Sequence-Model</b>  | <b>54</b> |
| 4.    | <b><i>Ergebnisse</i></b>   | <b>56</b> |
| 4.1.  | <b>Klinische Charakteristika der Patientenkohorte</b>                                    | <b>56</b> |
| 4.2.  | <b>Selektion der besten Features für den ML-Ansatz: Ausgewählte Features</b>             | <b>57</b> |
| 4.3.  | <b>Selektion des besten ML-Modells mit 5-facher Kreuzvalidierung</b>                     | <b>58</b> |
| 4.4.  | <b>Ergebnisse des SVM-Klassifikators mit Three-slice-Input</b>                           | <b>59</b> |
| 4.5.  | <b>Ergebnisse des SVM-Klassifikators mit Three-slice-Input am Validierungs-Datensatz</b> | <b>60</b> |
| 4.6.  | <b>Zonaler Ansatz: Ergebnisse</b>  | <b>61</b> |
| 4.7.  | <b>Deep Learning-Ansatz: Ergebnisse</b>  | <b>62</b> |
| 4.8.  | <b>Sequence-Model-Ansatz: Ergebnisse</b>   | <b>62</b> |
| 5.    | <b><i>Diskussion</i></b>   | <b>64</b> |
| 5.1.  | <b>Diskussion der Ergebnisse dieser Studie</b>   | <b>64</b> |
| 5.2.  | <b>Diskussion von Methoden und Limitationen dieser Studie</b>                            | <b>66</b> |
| 5.3.  | <b>Fazit und Ausblick</b>  | <b>68</b> |
| 6.    | <b><i>Zusammenfassung</i></b>  | <b>71</b> |
| 7.    | <b><i>Literaturverzeichnis</i></b>   | <b>73</b> |
| 8.    | <b><i>Lebenslauf</i></b>   | <b>91</b> |
| 9.    | <b><i>Danksagung</i></b>   | <b>93</b> |

# 1 ABKÜRZUNGSVERZEICHNIS

|         |   |
|---------|---|
| ACR     | American College of Radiology                                 |
| ADC     | Apparent Diffusion Coefficient                                |
| AFMS    | Anteriores fibromuskuläres Stroma                             |
| AI      | Artificial Intelligence                                       |
| AUC     | Area under the curve  |
| BI-RADS | Breast Imaging Reporting and Data System                      |
| BPH     | Benigne Prostatahyperplasie                                   |
| CAD     | Computerassistiertes Diagnosesystem /Computer-aided diagnosis |
| CAP     | College of American Pathologists                              |
| CART    | Classification and Regression Tree                            |
| CDS     | Computergestützte Diagnosesysteme                             |
| CIS     | Carcinoma in situ   |
| CLAIM   | Checklist for Artificial Intelligence in Medical Imaging      |
| CNN     | Convolutional Neural Network                                  |
| csPCa   | klinisch signifikantes PCa                                    |
| CT      | Computertomographie   |
| CV      | Cross-Validation  |
| DCE     | Dynamic contrast enhanced                                     |
| DGU     | Deutschen Gesellschaft für Urologie                           |
| DICOM   | Digital Imaging and Communications in Medicine                |
| DL      | Deep Learning   |
| DRU     | Digital-rektale Untersuchung                                  |
| DTPA    | Diethylenetriaminpentaessigsäure                              |
| DWI     | Diffusion-weighted imaging                                    |
| EAU     | European Association of Urology                               |
| EGFR    | Epidermal Growth Factor Receptor                              |
| ESUR    | European Society of Urology                                   |
| FDA     | Food and Drug Administration                                  |
| FLAIR   | fluid attenuated inversion recovery                           |
| Gd      | Gadolinium  |
| GLCM    | Gray-Level Co-Occurrence Matrix                               |
| GLDM    | Gray-Level Difference Matrix                                  |
| GLRLM   | Gray-Level Run Length Matrix                                  |
| GLSZM   | Gray-Level Size Zone Matrix                                   |
| GS      | Gleason-Score   |
| HGPIN   | High-Grade prostatiche intraepitheliale Neoplasie             |
| HOG     | Histogram of oriented gradients                               |
| IBSI    | Image Biomarker Standardisation Initiative                    |
| ISUP    | International Society of Urological Pathology                 |
| ITH     | Intratumorale Heterogenität                                   |
| KI      | Künstliche Intelligenz  |
| KM      | Kontrastmittel  |
| KNN     | Künstliches neuronales Netz                                   |
| LBP     | Local binary pattern  |
| LoG     | Laplacian of Gaussian   |
| LR      | Logistische Regression  |
| MHz     | Megahertz   |

---

|                 |   |
|-----------------|---|
| ML              | Machine Learning  |
| MONAI           | Medical Open Network for Artificial Intelligence  |
| mpMRT           | multiparametrische Magnetresonanztomographie  |
| mRMR            | Minimal-Redundancy-Maximal-Relevance  |
| MRT             | Magnetresonanztomographie   |
| non-csPCa       | klinisch nicht signifikantes PCa  |
| PCa             | Prostatakarzinom  |
| PI-RADS         | Prostate Imaging – Reporting and Data System  |
| PIA             | Proliferative inflammatorische Atrophie   |
| PIN             | Prostatische intraepitheliale Neoplasie   |
| PSA             | Prostata-spezifisches Antigen   |
| PSA-ACT         | Prostata-spezifisches Antigen und Alpha1-Antichymotrypsin-Komplex                               |
| PSAD            | Prostata-spezifisches Antigen-Dichte  |
| PZ              | Periphere Zone  |
| RF              | Random Forest   |
| RNA             | Ribonucleic acid (Ribonukleinsäure)   |
| ROC             | Receiver-Operating-Characteristic   |
| ROI             | Region of Interest  |
| RQS             | Radiomics Quality Scoring   |
| RSNA            | Radiological Society of North America   |
| SMOTE           | Synthetic Minority Over-Sampling Technique  |
| STR             | Society of Thoracic Radiology   |
| SVM             | Support Vector Machine  |
| T <sub>2w</sub> | T <sub>2</sub> -gewichtetes Bild  |
| TCIA            | The Cancer Imaging Archive  |
| TRIPOD          | Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis |
| TRUS            | Transrektaler Ultraschall   |
| TZ              | Transitionalzone  |
| UICC            | Union for International Cancer Control  |
| WHO             | World Health Organisation   |

## 2 EINLEITUNG UND GRUNDLAGEN

### 2.1. Epidemiologie des Prostatakarzinoms

Das Prostatakarzinom (PCa) ist die häufigste maligne Tumorerkrankung von Männern [1]. Pro Jahr werden weltweit 1,3 Millionen Männer neu mit PCa diagnostiziert [2]. Es stellt nach dem Bronchialkarzinom mit ca. 359.000 Todesfällen jährlich die zweithäufigste tumorbezogene Todesursache bei Männern dar [3].

In westlichen Industrieländern beträgt das Lebenszeitrisko für einen Mann, an einem PCa zu erkranken, etwa 40%. Allerdings wird nur ein Anteil von 10% der Erkrankungen symptomatisch, also klinisch signifikant, und das Mortalitätsrisiko beträgt etwa 2,5% [4, 5].

Laut Autopsiestudien kann bei einem von drei Männern über dem 50. Lebensjahr ein PCa festgestellt werden. In 80% der Fälle entsprechen diese latenten Karzinomen, also ohne klinische Symptomatik und mit einem geringen Grading [6, 7]. Damit ist das PCa eine Tumorerkrankung mit hoher Prävalenz bei gleichzeitig niedriger Mortalität.

Die Inzidenz des PCa steigt mit dem Alter [8]. Das mittlere Erkrankungsalter liegt bei 69 Jahren [9].

Das PCa ist eine der führenden Krebserkrankungen im Hinblick auf die Kosten für das Gesundheitssystem. Angesichts der demographischen Entwicklung kann von einer Zunahme von Prävalenz und Inzidenz und damit erhöhtem Versorgungsbedarf uroonkologischer Erkrankungen ausgegangen werden [10, 11]. Auch wegen der ansteigenden Kosten für neue Therapiemethoden wird das PCa somit eine zunehmende ökonomische Belastung für das Gesundheitssystem darstellen [10, 12-14].

### 2.2. Ätiologie und Risikofaktoren des Prostatakarzinoms

Das Lebensalter stellt neben der Ethnizität den wichtigsten Risikofaktor für das Auftreten eines PCa dar [15]. Demnach steigt das Risiko, an einem PCa zu erkranken, mit dem Lebensalter: Das Erkrankungsrisiko von 2,7% eines 55-jährigen Mannes erhöht sich im Alter von 65 Jahren auf 6,3% [11].

Die genauen ursächlichen Mechanismen, die der Entstehung des PCa ursächlich zugrunde liegen, sind nach wie vor unbekannt. Erwiesen ist aber laut einigen Studien, dass bei der Karzinogenese des PCa Genveränderungen eine Rolle spielen [16-18]. In einer Studie von Su et al. konnten Genmutationen identifiziert werden, die ursächlich für die Entstehung von PCa sein können. Außerdem konnte in der Studie nachgewiesen werden, dass die einzelnen Herde eines multifokalen PCa unabhängig voneinander entstehen können [19].

Entsprechend ist auch die familiäre Prädisposition gewichtig: Verglichen mit anderen Karzinomen liegt beim PCa am häufigsten eine familiäre Vorbelastung vor. So lässt sich in einem Anteil von rund 20% eine positive Familienanamnese feststellen [20, 21]. Erkrankt ein Verwandter ersten Grades an PCa, so verdoppelt sich das Erkrankungsrisiko; sind zwei Verwandte ersten Grades betroffen, steigt das Erkrankungsrisiko um das 5- bis 11-fache [22]. Darüber hinaus sinkt im Falle einer positiven Familienanamnese das Erkrankungsalter [23].

Neben der Genetik werden exogene Einflüsse wie Ernährungsweise, Lebensort und sozioökonomische Gegebenheiten angenommen. Es besteht zum einen ein klares West-Ost- sowie Nord-Süd-Gefälle für die Inzidenz von PCa. So ist in westlichen Ländern die Prävalenz

des PCa deutlich höher als im asiatischen Raum [24]. Zum anderen konnte nachgewiesen werden, dass Männer mit demselben genetischen Hintergrund dasjenige Risiko für ein PCa aufweisen, das in der Region ihres Wohnsitzes vorherrscht: Für asiatische Männer ist das Erkrankungsrisiko im Vergleich zu afroamerikanischen oder kaukasischen Männern geringer. Eine Verlagerung des Lebensmittelpunktes von Männern mit asiatischer Herkunft in westliche Länder geht mit einem erhöhten Risiko für ein klinisch signifikantes PCa einher: Im Falle einer Standortverlagerung eines asiatischen Mannes nach Kalifornien gleicht sich dessen Erkrankungsrisiko dem des amerikanischen Mannes an [15, 25-27].

Der Testosteronspiegel wird als möglicher Risikofaktor für PCa diskutiert, da bekannt ist, dass Initiation und Progression des PCa dem Einfluss von Androgenen unterliegen können [28]. Bei Eunuchen wurden bislang keine PCa beschrieben, und die medikamentöse Androgensuppression bzw. funktionelle Kastration ist eine effektive First-Line-Therapie bei PCa in fortgeschrittenem Stadium [29].

### 2.3. Anatomie des Prostatakarzinoms

Die Prostata ist eine akzessorische Geschlechtsdrüse des Mannes und liegt im Becken zwischen Harnblase und Beckenboden. Die Basis prostatae ist mit dem Boden der Harnblase verwachsen, der Apex liegt dem Beckenboden auf.

Sie lässt sich in vier verschiedene Zonen einteilen, die sich zwiebelschalenartig um die Harnröhre legen, welche als Leitstruktur dient. Die periphere Zone (PZ), welche mit ca. 70 % den größten Teil des prostatistischen Drüsengewebes ausmacht, umgibt dorsal, apikal und lateral die Transitionalzone (TZ). Diese besteht aus zwei Lappen, die zwischen proximaler Urethra und PZ liegen. Die zentrale Zone ist kegelförmig und reicht vom Blasenansatz bis zum Colliculus seminalis in der Pars prostatica der Urethra. Die Oberfläche der Vorderseite der Prostata besteht aus dem anterioren fibromuskulären Stroma (AFMS), welches kein Drüsengewebe enthält, aber fließend in dieses übergeht. Abb. 1 veranschaulicht die Anatomie der Prostata.

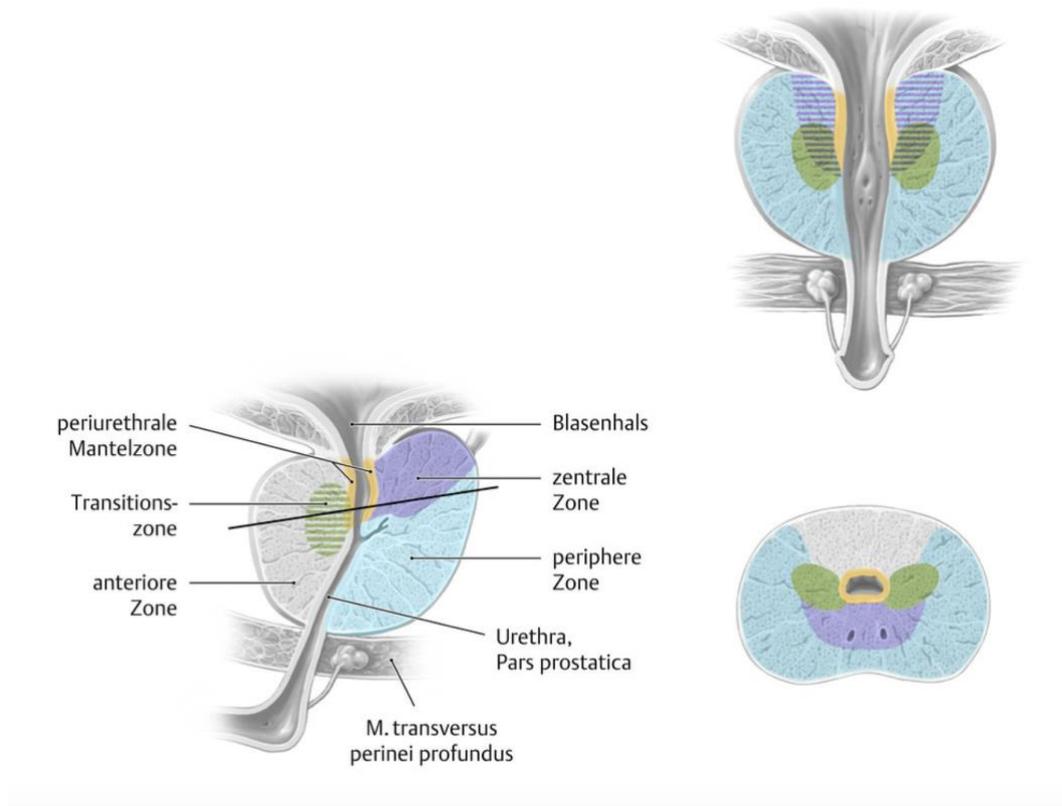


Abb. 1.: Darstellung der Anatomie der Prostata. Koronare (oben rechts), sagittale (unten links) und axiale Schnittführung (unten rechts) [30]. Mit freundlicher Genehmigung von Thieme.

Etwa 80-85% der PCa befinden sich in der peripheren Zone (PZ) und entstehen multizentrisch [31]. 15-20% der PCa entstehen in der Transitionalzone (TZ), während Karzinome der zentralen Zone selten sind [32].

#### 2.4. Histopathologie des Prostatakarzinoms

Das PCa ist in 95% der Fälle als Adenokarzinom zu klassifizieren. Seltener sind Urothelkarzinome mit einem Anteil von 3%, noch seltener sind Plattenepithel- und Basalzellkarzinome [33]. In der WHO-Klassifikation (2016) werden darüber hinaus das pseudohyperplastische, muzinöse, sarkomatoide, atrophe, siegelringzellige und kleinzellige PCa als Beispiele für besonders seltene Entitäten genannt [34].

Die maligne Transformation beginnt mit der prostatichen intraepithelialen Neoplasie (PIN), welche eine Übergangsform bzw. Präkanzerose des PCa darstellt. Sie entspricht dem Carcinoma in situ (CIS). Diese kann in eine „Low-Grade“-PIN und „High-Grade“-PIN (HGPN) unterschieden werden, wobei jedoch nur die HGPN als fakultative Vorläuferläsion des PCa gilt [33, 35]. Insbesondere bei multifokalem Vorkommen in Stanzbiopsien wird die HGPN als Indikatorläsion der PZ gewertet [36]. Die PIN wiederum wird vermutlich durch eine chronische Entzündung der Prostata begünstigt, der proliferativen inflammatorischen Atrophie (PIA) [37, 38].

Auf die Transformation vom epithelialen zum mesenchymalen Phänotyp folgt die Gefäßinvasion, wodurch entartete Zellen in den Körperkreislauf gelangen. Dies ermöglicht

anschließend eine Extravasation von Tumorzellen aus dem Gefäßsystem, daraufhin einen Organbefall und somit die Entstehung von Fernmetastasen [39].

Das PCa weist in Bezug auf seine Histologie oftmals eine hohe intratumorale Heterogenität (ITH) auf, was sich entscheidend auf den Therapieerfolg auswirken kann [40, 41].

#### 2.4.1. Gleason-Score

Die histologische Beurteilung der Drüsenmorphologie von PCa ist Grundlage für den Gleason-Score (GS). Als einer der wichtigsten Prognosefaktoren in der Prostatakrebsdiagnostik klassifiziert er den Grad der Entdifferenzierung und damit die Tumoraggressivität [42].

1966 entwickelte der amerikanische Pathologe Donald F. Gleason den Score, welcher das Zellwachstumsmuster in Gewebeproben der Prostata in Grade von 1 – 5 unterteilt.

Gewebe aus einer Stanzbiopsie oder einem Prostatektomiepräparat wird histologisch untersucht. Dabei wird ausschließlich die Architekturstörung des Drüsengewebes beurteilt, Kernatypien spielen keine Rolle [43].

Je stärker das untersuchte Gewebe von der physiologischen Drüsenarchitektur abweicht, also je weniger differenziert das Gewebe ist, desto höher fällt der ermittelte Punktwert aus. Im Jahr 2000 wurde der GS von Gewebe aus einer Stanzbiopsie in der Leitlinie des College of American Pathologists (CAP) als „the most & the worst“ definiert [44]. Bei der Stanzbiopsie wird demnach der Punktwert des am häufigsten vorkommenden und des am schlechtesten differenziertesten Gewebes addiert. So kann der GS einen Wert zwischen 2 und 10 ergeben. Je höher der Punktwert, desto aggressiver der Tumor und desto ungünstiger ist die Prognose. PCa mit einem GS von  $\geq 7$  oder einem überwiegenden Gleason-Grad  $\geq 4$  gelten als high-grade PCa [45].

Im Fall einer Prostatovesikulektomie wird der GS aus der Summe der beiden dominantesten Zellwachstumsmuster gebildet, sowie der höchste Grad als tertiäres Muster angegeben [46].

Im pathologischen Befund werden üblicherweise zusätzlich zur Summe auch die einzelnen Gleason-Werte in der Reihe ihrer Häufigkeit angegeben, sodass die Graduierung des häufigsten Zelltyps deutlich wird, z.B.: Gleason 8 (5+3). Eine Studie konnte belegen, dass im Falle eines Scores von 7 der dominante Grad entscheidend für Therapie und Prognose ist. Daher wird oft gesondert durch „a“ oder „b“ angegeben, welcher Gleason-Grad dominant ist, z.B. bei dominantem Grad 3: 7a; bei dominantem Grad 4: 7b (4+3) [47, 48]. Der GS korreliert darüber hinaus mit der Überlebensrate nach radikaler Prostatektomie [49].

Tab. 1.: Gleason-Score-Grade und morphologische Kriterien nach Kristiansen et al. [50]

| Gleason-Grad | Morphologische Kriterien   |
|--------------|--|
| 1            | „Umschriebene Knoten dicht gelagerter, scharf begrenzter Drüsen: rundlich bis oval, mittelgroß (größere Drüsen als in Muster 3). In Karzinomen kaum vorkommend“  |
| 2            | „Ähnlich Muster 1; noch weitgehend umschriebene, z. T. nicht ganz scharf begrenzte Tumorknoten mit minimaler Stromainfiltration. Aufgelockertes Drüsenarrangement. In Karzinomen kaum vorkommend.“         |
| 3            | „Diskrete, gut umschriebene Drüsen in infiltrativer Lagerung zwischen nichtneoplastischen Drüsen. Variation in Form und Größe.“  |
| 4            | „Fusionierte mikroazinäre Drüsen; wenig differenzierte Drüsen mit schwach ausgebildeten Drüsenlichtungen, alle kribriformen Drüsen einschließlich glomeruloider Formen, auch sog. hypernephroides Muster.“ |
| 5            | „Verlust drüsiger Strukturen, stattdessen solide Epithelstränge, komplexe oder Einzelzellen. Zentrale Nekrosen in Drüsen, entsprechend einem Komedokarzinom.“  |

Basierend auf dem GS wurde durch die International Society of Urological Pathology (ISUP) eine neue Grad-Gruppeneinteilung eingeführt. Dieses ISUP-Graduierungssystem wurde 2016 in die WHO-Klassifikation für „Harnwegssystem und männliche Genitalorgane“ aufgenommen [51]. Sie unterteilt fünf prognostische Gruppen, für die in einer Studie signifikante Unterschiede des biochemisch Rezurrenz-freien 5-Jahresüberlebens erwiesen wurden [52].

## 2.5. Klinik des Prostatakarzinoms

Da die bevorzugte Lokalisation des PCa in der peripheren und damit äußeren Zone liegt, bleibt die Erkrankung meist lange asymptomatisch. Symptome wie Harnverhalt, Inkontinenz, Hämaturie und Impotenz treten erst ab lokal fortgeschrittenen Stadien auf. Im Zuge der intraprostatichen Ausbreitung erfolgt erst spät die Infiltration des periurethralen Drüsenbereichs. Nach der Kapselüberschreitung werden Nachbarorgane wie Harnblase und Samenblasen infiltriert [53].

Entsprechend werden zum Zeitpunkt der Diagnosestellung bei 50% der Patienten ein fortgeschrittenes Tumorstadium und bei bis zu 10% der Patienten bereits Knochenmetastasen diagnostiziert [35].

Das PCa metastasiert auf sowohl lymphogenem als auch hämatogenem Wege, sodass neben iliakalen, retroperitonealen und paraaortalen Lymphknoten auch Leber und Lunge befallen werden [35, 54]. Die fortgeschrittene Tumorprogression des PCa manifestiert sich häufig mit osteoplastischen Metastasen und dadurch bedingten Knochenschmerzen [55].

## 2.6. Klassifikation des Prostatakarzinoms

PCa werden in der TNM-Klassifikation nach der Ausdehnung des Primarius (T-Stadium; 1 – 4, zusätzlich durch Buchstaben unterteilt), dem Befallsstatus regionärer Lymphknoten (N-Stadium; „0“ = nicht vorhanden, „1“ = vorhanden) sowie dem Vorliegen von Fernmetastasen (M-Stadium; analog N, zusätzlich a - c) eingeteilt. Basiert das Staging auf klinischer Diagnostik

(Prostata-spezifisches Antigen (PSA), digital-rektale Untersuchung (DRU), bildgebende Verfahren), wird dies durch ein „c“ vor der TNM-Klassifikation kenntlich gemacht. Ein „p“ kennzeichnet das Staging anhand des endgültigen histopathologischen Befundes des Prostatektomiepräparats [11].

Nach der Union for International Cancer Control (UICC) werden die Stadien T1-T2 N0, M0 als „lokal begrenztes PCa“ und T3-T4 N1 und/oder M1 als „fortgeschritten/metastasiert“ subsumiert [56, 57].

So wird eine Stadieneinteilung definiert, die als Grundlage für klinische Studien und das weitere klinische Management dient.

Eine Einteilung in Risikogruppen anhand objektiver Parameter ermöglicht die Auswahl der bestmöglichen, stadienspezifischen Therapiestrategie. Außerdem kann somit eine Einschätzung des Rezidivrisikos getroffen werden. Nach D’Amico werden lokal begrenzte PCa anhand von PSA-Wert (die Erläuterung hierzu folgt im Abschnitt „Früherkennung und Prostata-spezifisches Antigen“), GS, und TNM-Stadium in niedriges, intermediäres und hohes Risiko kategorisiert [57, 58]:

- Niedriges Risiko:  $PSA \leq 10$  ng/ml,  $GS \leq 6$ , cT1c/2a
- Intermediäres Risiko:  $PSA 10 - 20$  ng/ml,  $GS = 7$  oder cT2b
- Hohes Risiko:  $PSA > 20$  ng/ml oder  $GS \geq 8$  oder cT2c

## 2.7. Diagnostik des Prostatakarzinoms

Eine kurative Therapie des PCa ist nur im lokalisierten bzw. organbegrenzten Stadium möglich [11]. Für das PCa in seinem Frühstadium ist eine gute Therapierbarkeit und ein exzellentes tumorspezifisches Überleben zu erwarten. Entsprechend bedeutsam sind die frühe Detektion, eine möglichst genaue Lokalisierung, sowie die Charakterisierung und Risikostratifizierung des PCa. [59].

### 2.7.1. Früherkennung: Prostata-spezifisches Antigen und digital-rektale Untersuchung

Im Rahmen der Früherkennung von PCa wird vornehmlich der Serumwert des Glykoproteins Prostata-spezifisches Antigen (PSA-Wert) bestimmt.

Dem PSA-Wert als laborchemischer Parameter kommt in der Diagnostik und Verlaufskontrolle des PCa eine große Bedeutung zu. Das PSA ist eine ausschließlich von Prostatadrüsen synthetisierte Serinprotease, die von der gesunden Prostata physiologischerweise nur in geringem Maße sezerniert wird.

Der Nutzen des PCa-Screenings ist zwar umstritten [60]. Seit der Zulassung des PSA-Screenings 1986 konnte jedoch eine Erhöhung des diagnostizierten Anteils von kleinen, nicht tastbaren Karzinomen in Frühstadien verzeichnet werden [6, 61, 62].

Männer ohne Risikofaktoren sollen laut der Leitlinie der Deutschen Gesellschaft für Urologie (DGU) ab ihrem 45. Lebensjahr und einer Lebenserwartung von  $> 10$  Jahren über die Möglichkeit zum Screening informiert werden, bei vorliegenden Risikofaktoren bereits ab einem Alter von 40 Jahren. Abhängig vom Alter des Patienten und initialem PSA-Wert werden Verlaufskontrollen in jährlichen ( $PSA > 2$  ng/ml) bis vierjährigen Intervallen (40 – 50 Jahre,  $PSA < 1$  ng/ml) empfohlen. Neben der Bestimmung des Serum-PSA-Werts wird im Rahmen der Früherkennung auch die digital-rektale Untersuchung (DRU) als klinisches

Untersuchungsverfahren durchgeführt. Sie erlaubt durch Palpation eine erste Einschätzung des Prostatavolumens, sowie die Beurteilung von Schleimhautverschieblichkeit, Konsistenz und Oberflächenveränderungen (z.B. Verhärtungen) der Prostata [11].

Während die DRU alleine aufgrund niedrigerer Sensitivität und Spezifität nicht als ausreichend zur Diagnose von PCa gilt, können Spezifität und Sensitivität des PSA-Wertes zur Erkennung von lokalisiertem PCa durch eine Kombination mit DRU verbessert werden [63, 64].

Die Spezifität des PSA-Wertes kann außerdem erhöht werden, indem der Quotient von freiem bzw. Gesamt-PSA bestimmt wird. Der Anteil an im Komplex (PSA und Alpha1-Antichymotrypsin, PSA-ACT) gebundenen PSA im Serum ist bei PCa höher und der frei vorliegende Anteil niedriger als bei der benignen Prostatahyperplasie (BPH). Ein Größenanteil des freien PSA von unter 15% weist auf ein PCa hin. Die Bestimmung der PSA-Ratio wird besonders bei Patienten mit PSA-Werten im „Grauzonen-Bereich“ (2,4-10 ng/ml) zu Rate gezogen [65, 66].

Die PSA-Dichte (PSAD) ist ein Wert, der bei Vorliegen eines PCa höher ausfällt als bei gutartigen Hyperplasien. Er wird aus PSA-Wert in ng/ml und Prostatavolumen in ml errechnet. Die Volumetrie der Prostata erfolgt mittels transrektalen Ultraschalls (TRUS).

Laut einer Übersichtsarbeit von Mistry und Cable beträgt der Anteil der durch PSA/DRU-Screening im Frühstadium erkannten, lokalisierten PCa rund 84% [63]. Nach Roobol et al. senkt das PSA-Screening das Risiko gescreenter Männer der Altersgruppe von 55 – 69 Jahren an PCa zu versterben um 31% [67, 68].

Bildgebende Verfahren werden als primäre Untersuchungsmethode zur Früherkennung nicht empfohlen [11].

### 2.7.2. Transrektaler Ultraschall

Der transrektale Ultraschall (TRUS) wird mit einer rektal eingeführten Ultraschallsonde mit Frequenzen von 7,5 bis 10 MHz durchgeführt.

Er besitzt allerdings nur eine niedrige Sensitivität und Spezifität im Hinblick auf PCa (26,9% – 33% respektive 88%; der positiv prädiktive Wert liegt bei 57%) [69-71].

Gemäß der S3-Leitlinie zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms der Deutschen Krebsgesellschaft wird der TRUS unterstützend zur Stanzbiopsie eingesetzt, wo er zur gezielten Probengewinnung unter Sichtkontrolle dient. Darüber hinaus spielt er jedoch in der Früherkennung nur eine untergeordnete Rolle [11, 72].

### 2.7.3. Biopsie

Im Rahmen der Früherkennung wird bei einem PSA-Wert ab 4 ng/ml, einem karzinomverdächtigen DRU-Befund oder einem auffälligen PSA-Anstieg in der S3-Leitlinie zum Prostatakarzinom eine stanzbiopsische Abklärung empfohlen. Zur Indikationsstellung sollen Alter und Lebenserwartung, Familienanamnese, Komorbiditäten und etwaige vorangegangene Biopsien mitbetrachtet werden [11, 66].

Von der European Association of Urology (EAU) wird zur stanzbiopsischen Abklärung eine systematische, symmetrische Probenentnahme von 10 - 12 Gewebezylindern empfohlen [73, 74].

Dazu wird die Prostata schematisch in Sextanten unterteilt und unter Antibiotikaphylaxe zweimal pro Sextant biopsiert.

So soll eine möglichst gleichmäßige Verteilung der biopsierten Areale bei möglichst niedriger Komplikationsrate (z.B. Infektion, Bakteriurie) sichergestellt werden [75].

Außerdem kann eine Biopsie suspekter Areale unter digitaler bzw. palpatorischer Steuerung oder TRUS-gesteuert erfolgen. Bei der zielgerichteten Biopsie wird vor der Biopsie eine MRT der Prostata erstellt und suspekte Läsionen identifiziert, die anschließend gezielt biopsiert werden.

Laut der aktuellen Studienlage entgehen jedoch beiden Methoden eine substanzielle Anzahl an klinisch signifikanten PCa [76]. Die systematische Biopsie detektierte beispielsweise in einer großen prospektiven Studie bei einem Anteil von 16% der Patienten klinisch signifikante PCa, bei denen in der multiparametrischen Magnetresonanztomographie (mpMRT) keine suspekten Läsionen nachweisbar war [77].

Laut einigen Studien ist die zielgerichtete Biopsie der systematischen Biopsie zur Detektion von klinisch signifikanten bzw. high risk PCa überlegen [77-80]. Laut Siddiqui et al. und Schoots et al. detektiert sie nicht nur mehr klinisch signifikante PCa, sondern auch weniger klinisch nicht signifikante PCa [81, 82]. In anderen Studien hingegen zeigte sich die zielgerichtete Biopsie der systematischen Biopsie gegenüber gleichwertig [76, 83, 84]. Einig ist sich der Großteil der Studien allerdings darüber, dass die beste Detektionsrate für klinisch signifikante PCa durch eine Kombination aus systematischer TRUS-gesteuerter und zielgerichteter Biopsie (MRT-3D-TRUS-Fusionsbiopsie) erreicht wird. Die MRT-3D-TRUS-Fusionsbiopsie ist ein Verfahren, bei dem die zuerst angefertigten mpMRT-Daten mit einem dreidimensionalen Ultraschallbild computergestützt kombiniert bzw. fusioniert werden. Somit lassen sich karzinomsuspekte Areale zielgerichtet biopsieren, was die Diagnosesicherheit erhöht [76, 77, 83, 84].

Die aktuelle interdisziplinäre S3-Leitlinie empfiehlt nach negativer systematischer Biopsie im Rahmen der Primärdiagnostik des PCa eine MRT-Bildgebung und ggf. eine kombinierte gezielte und systematische Biopsie [11].

#### 2.7.4. Multiparametrische MRT der Prostata

Das Verfahren der Wahl zur Diagnostik bzw. Abklärung von PCa ist insbesondere bei Patienten mit negativer Biopsie und/oder erhöhten PSA-Werten die mpMRT der Prostata.

Sie bietet einen hervorragenden Weichteilkontrast, liefert in einer Untersuchung sowohl morphologische als auch funktionelle Informationen, und lässt eine gute Beurteilung benignen oder malignen Veränderungen der Prostata zu [85]. Als nicht-invasive, nebenwirkungsarme, detaillierte und dreidimensionale Karzinombildgebung bietet sie erhebliche Vorteile gegenüber anderen Untersuchungsmethoden. Aufgrund unterschiedlicher Signalcharakteristika lassen sich die vier Drüsenabschnitte der Prostata MRT-morphologisch in die Bereiche der zentralen und peripheren Drüse zusammenfassen [86]. Dabei gehören die anatomische zentrale Zone, die Transitionalzone und das peripher urethrale Drüsengewebe zur zentralen Drüse [87].

Die MRT erlaubt verschiedene Bildaufnahmetechniken bzw. Sequenzen, die unterschiedliche Signalintensitäten und Bildkontraste in der Auslesung des Kernspinresonanzsignals generieren und somit unterschiedliche Informationen liefern [88, 89].

Die mpMRT der Prostata setzt sich aus einer hochauflösenden T<sub>2</sub>-gewichteten Sequenz (T<sub>2</sub>w) und den beiden funktionellen Bildgebungsmethoden der diffusionsgewichteten Bildgebung („diffusion-weighted imaging“, DWI) sowie der dynamischen kontrastverstärkten Bildgebung („dynamic contrast enhanced MR“, DCE-MR) zusammen [90, 91].

Der mpMRT kommt in der Prostatakarzinomdiagnostik eine zentrale Bedeutung zu und soll nach genau definierten Qualitätsstandards durchgeführt werden. Durch die verschiedenen Bildgebungstechniken (T<sub>2</sub>w, DCE-MRT, DWI) haben sich die Anwendungsbereiche der

mpMRT erweitert, und sie spielt nicht nur in der Tumordetektion, sondern auch in Staging, Active Surveillance, Therapiemonitoring und Follow-ups eine substantielle Rolle [92].

Die einzelnen Bildgebungssequenzen weisen zur Detektion von PCa eine unterschiedliche Performance auf. Das PCa kann in keiner der einzelnen Sequenzen alleine ausreichend verlässlich diagnostiziert und charakterisiert werden [93]. Die kombinierte Betrachtung der verschiedenen Sequenzen hingegen liefert eine bessere Detektionsrate für PCa [94].

Im Folgenden werden die einzelnen Sequenzen kurz erläutert.

#### 2.7.4.1. T<sub>2</sub>-gewichtete Sequenz

Die T<sub>2</sub>-gewichtete Sequenz liefert hochwertige anatomische Informationen: Die zonale Anatomie und Organkapsel der Prostata lassen sich in der T<sub>2</sub>w gut voneinander differenzieren. Somit ist sie geeignet, um Läsionen im Prostatagewebe zu identifizieren. In der Prostate Imaging – Reporting and Data System Version 2 - Klassifikation (PI-RADS v2) (die Erläuterung der Klassifikation folgt im Abschnitt „Prostate Imaging – Reporting and Data System“) ist sie für die Detektion von PCa in der zentralen Zone die Leitsequenz [95].

Diagnostisches Kriterium für PCa in der T<sub>2</sub>w ist eine oft rundliche, tropfenförmige, irregulär begrenzte Hypointensität. Jedoch ist dies nicht immer spezifisch für PCa, sondern auch Veränderungen wie chronische Prostatitis, Narben, Atrophien, Hyperplasien, postbiotische Blutung und postradiogene Veränderungen können sich ähnlich wie Tumorgewebe darstellen [96].

Auch können Kapselgrenzen, Samenblasen, und neurovaskuläre Bündel mit der T<sub>2</sub>w-Sequenz auf extrakapsuläre Ausbreitung hin untersucht werden. Diese beeinflusst als ein entscheidender Prognosefaktor die Therapieentscheidung maßgeblich, da eine extrakapsuläre Ausbreitung durchaus in Frühstadien von PCa vorkommen kann und da sie die radikale Prostatektomie als kurative Therapie ausschließt [73]. Laut einer Studie von [97] tritt die extrakapsuläre Ausbreitung mit einer Inzidenz von 60% bereits bei Patienten mit einem klinischen T1/T2 Stadium auf.

Die Studienlage zur diagnostischen Genauigkeit von T<sub>2</sub>w-Sequenzen zur Beurteilung von extrakapsulärer Ausbreitung ist durchwachsen und rangiert je nach Erfahrung der befundenden Ärzt:innen von 50% bis 92% [98, 99]. Isoliert betrachtet liegt die Detektionsrate der T<sub>2</sub>w-Sequenz für PCa bei 50-63% [100, 101]. Die Sensitivität und Spezifität werden durch die kombinierte Befundung von T<sub>2</sub>w, DWI und DCE-MRI verbessert [94].

#### 2.7.4.2. Diffusionsgewichtete Bildgebung

Die diffusionsgewichtete Bildgebung (engl.: diffusion weighted imaging, DWI) wird seit den 1990ern zur Detektion von PCa verwendet [95, 102]. Sie erstellt den Bildkontrast anhand der Darstellung der zufälligen mikroskopischen Bewegungen von Flüssigkeiten in biologischen Geweben, der Brown'schen Molekularbewegung. Sie wird von Makromolekülen, Zelldichteunterschieden, Zellmembranen, Größe des Extrazellulärraumes bzw. des extravaskulären Raumes beeinflusst und unterschiedlich stark eingeschränkt. Diese sogenannte Diffusionsrestriktion bzw. Restriktion der mittleren Diffusivität lässt sich mit der Mikrostruktur eines Gewebes in Beziehung setzen; so weist z.B. Tumorgewebe normalerweise eine höhere Diffusionsrestriktion auf als gesundes Gewebe [103].

Maligne Veränderungen der Prostata stellen sich in der DWI mit hyperintensem Signal dar [104].

Die DWI ist für die Detektion von PCa in der PZ gut geeignet [105-107]. In der PI-RADS v2.1. ist sie bei der Befundung die Leitsequenz zur Detektion von PCa in der peripheren Zone [91]. Die DWI Daten lassen sich zusätzlich zu den aufgenommenen DWI-Bildern auch als Apparent Diffusion Coefficient (ADC) -Bilder beurteilen: Basierend auf mindestens zwei DWI Bildern mit unterschiedlichen Diffusionswichtungsfaktoren ( $b$ -Werten) können diese sogenannten „ADC-maps“ errechnet werden, die die Diffusion von Wassermolekülen quantifizieren. Diese Größe wird in Quadratmillimeter pro Sekunde ( $\text{mm}^2/\text{s}$ ) angegeben und ist ein statistisches Maß für die Fläche, die die Wassermoleküle in einer Sekunde überquert haben. Dieser Wert wird für jedes Bild- bzw. Volumenelement (Pixel/Voxel) berechnet und als ADC-map dargestellt. Das PCa stellt sich in der ADC hypointens dar [104, 108]: Maligne Veränderungen haben eine niedrigere mittlere Diffusivität, ergo einen niedrigeren mittleren ADC-Wert („meanADC“). Jedoch ist der ermittelte ADC-Wert stark von den gewählten  $b$ -Werten der zugrunde liegenden diffusionsgewichteten Aufnahmen abhängig.

Während bei niedrigen  $b$ -Werten neben Diffusions- auch Perfusionseffekte die Signalintensität beeinflussen können, dominiert bei hohen  $b$ -Werten ( $> 1000 \text{ s/mm}^2$ ) der Einfluss der Diffusion. Laut Wang et al. ergibt ein  $b$ -Wert von  $1500 \text{ s/mm}^2$  die höchste Effizienz in der Detektion von PCa [109].

In der PI-RADS v2.1. wird für die Erstellung der DWI-Sequenzen die Verwendung eines niedrigen  $b$ -Werts (ca.  $0 - 100 \text{ s/mm}^2$ ) und eines mittleren  $b$ -Werts ( $800 - 1000 \text{ s/mm}^2$ ) empfohlen, zusätzlich zum bedeutenden hohen  $b$ -Wert ( $\geq 1400 \text{ s/mm}^2$ ), welcher entweder interpoliert oder separat akquiriert werden kann [110].

Laut einigen Studien kann der ADC-Wert mit der Tumoraggressivität korreliert werden: Niedrige ADC-Werte korrelieren mit höheren Gleason Scores [111-116].

Eine Studie, in der die Histopathologie mit den diagnostischen Kriterien in  $T_2w$  und der Diffusionsrestriktion mit niedrigem ADC-Wert korreliert wurde, konnte eine Spezifität und PPV von 100% zeigen [117].

Die ADC kann darüber hinaus laut Park et al. mit einer Sensitivität von 91% und einer Spezifität von 100% zwischen TZ-PCa und BPH differenzieren [118].

Die Kombination von DWI mit  $T_2w$  ( $\text{AUC} = 0,93$ ) verbessert signifikant die Detektion von PCa gegenüber der konventionellen  $T_2w$ -Bildgebung ( $\text{AUC} = 0,87$ ) [119, 120].

#### 2.7.4.3. Dynamische kontrastmittelbasierte Perfusionsbildgebung

Mit der DCE-MRT können Regionen mit vermehrter Gewebepfusion bzw. Tumorneovaskularisation visualisiert werden.

In der MRT im Allgemeinen und zur Diagnostik des PCa im Speziellen wird aufgrund guter Verträglichkeit, schneller renaler Ausscheidung und hoher Relaxivität gadoliniumhaltiges (Gd) Kontrastmittel (KM) verwendet, das aufgrund seiner Toxizität in einem Chelat an Diethyltriaminpentaessigsäure (DTPA) gebunden appliziert wird (Gd-DTPA) [121, 122].

Die pharmakokinetischen Eigenschaften des KM werden durch die erhobenen messtechnisch-quantitativen und morphologischen Parameter reflektiert.

Nach intravenöser Applikation des KM kann dieses nach nur wenigen Sekunden im interstitiellen Raum nachgewiesen werden. Anhand des KM-Signalverlaufs kann mit schnellen MR-Sequenzen die Mikrozirkulation bzw. Perfusion visualisiert und beurteilt werden [123]. Die Unterschiede der Signal-Intensity-Time-Verläufe von benignem und malignem Gewebe können qualitativ, semi-quantitativ oder quantitativ dargestellt werden.

Das PCa kontrastiert sich dem Normalgewebe gegenüber typischerweise durch einen starken, früharteriellen Signalintensitätsanstieg („Enhancement“), und stellt sich mit einer schnellen Abflutung („Washout“) dar, was sich auf die Tumorneovaskularisation und damit lokal verstärkte Durchblutung zurückführen lässt [124, 125]. Dies ergibt zusammen mit der maximalen KM-Aufnahme („Peak“) und deren zeitlichem Verlauf („Time to Peak“) eine dynamische Kurve pro Voxel.

Außerdem ist die Permeabilität der zu- und abführenden Gefäße im Tumor erhöht, was eine Metastasierung begünstigt und sich mit der schnellen Extravasation des KM in den Extrazellularraum („leakage space“) darstellt [95, 126]. Die Prognose von PCa verschlechtert sich entsprechend mit steigender Anzahl der pathologischen Gefäße im PCa [127, 128].

Neben der Einschätzung der Tumorperfusion und der Gefäßpermeabilität kann auch eine Kapselinfiltration oder eine extrakapsuläre Ausbreitung mit der DCE-MR mit großer diagnostischer Sicherheit detektiert werden, worin sie der ADC und der T<sub>2</sub>w- Bildgebung überlegen ist [85, 129].

In der Detektion von PCa im Frühstadium und Staging weist die DCE-MR eine hohe Sensitivität und Spezifität auf und kann sogar bis zu einem gewissen Grad die vorliegende Histopathologie vorhersagen [122]. Laut einer Studie von Taghipour et al. verbessert das Hinzuziehen der DCE die Diagnosegenauigkeit von klinisch signifikanten PZ-lokalisierten PCa, die anhand der DWI in der PI-RADS v2 nicht eindeutig klassifiziert wurden, und auch Ullrich et al. wiesen eine verbesserte Detektionsrate von PI-RADS-4 Patienten sowie vermindertes Undergrading von peripheren PCa nach [130, 131]. Die Rolle von DCE-MR für die PI-RADS v2.1 ist dennoch kontrovers diskutiert [132]. Sie wird in der PI-RADS v2 nur zur Beurteilung von unklaren Befunden in der PZ herangezogen [91].

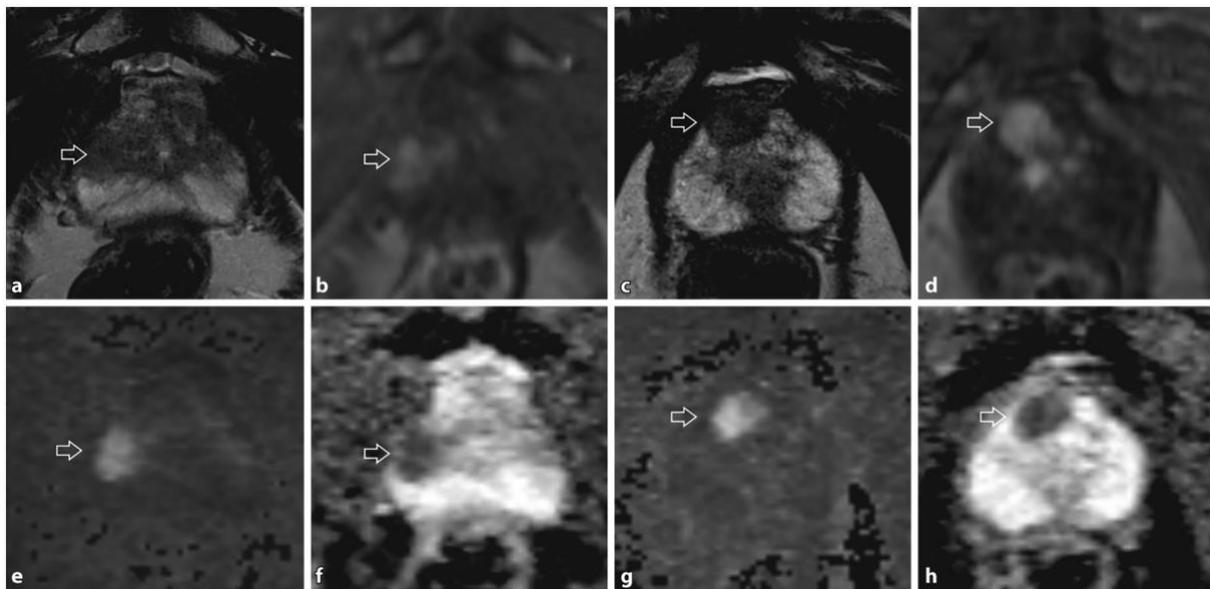


Abb. 2: Beispielhafte mpMRT-Aufnahmen der Prostata. Die Pfeile markieren ein laterales PCa mit GS 3 + 3 (a, b, e, f) und ein ventrales PCa der PZ mit GS 4 + 4 (c, d, g, h). Aufnahmemodalitäten: axiale T<sub>2</sub>w (a, c), DCE (b, d), DWI (e, g), ADC-map (f, h). Hier zeigen sich die im Text beschriebenen bildgebenden Eigenschaften des PCa [133].

Mit freundlicher Genehmigung durch: Springer Nature / Der Radiologe 60, 48-55 [Artificial intelligence and radiomics in MRI-based prostate diagnostics]. Hamm, C.A., Beetz, N.L., Savic, L.J., and Penzkofer, T. (2020).

### 2.7.5. Prostate Imaging – Reporting and Data System

Um die Diagnosegenauigkeit für PCa zu verbessern, die Kriterien und Interpretation zu objektivieren und die Interreader-Variabilität bei der Begutachtung von mpMRT zu optimieren, wurde 2012 von der European Society of Urogenital Radiology das Prostate Imaging – Reporting and Data System eingeführt. Es wurde in Anlehnung an das in der Mammadiagnostik bereits fest etablierte Befundungsschema „Breast Imaging Reporting and Data System (BI-RADS)“ entwickelt und ist ein breit eingesetztes System zur Befunderstellung.

PI-RADS gibt Mindeststandards für technische Parameter für die mpMRT der Prostata vor und ermöglicht die einheitliche Klassifikation und Kommunikation von suspekten Läsionen [91]. Definierte Kriterien für das Erscheinungsbild von Läsionen werden in einen Punktescore auf einer Likert-Skala von 1 bis 5 übertragen (vgl. Tab. 2) und somit suspekta Befunde klassifiziert. Um einen möglichst einfach zu kommunizierenden Score unabhängig von den verwendeten Methoden zu erhalten, wird ein Summenscore aus den einzelnen Punktwerten gebildet, und dieser in einen Gesamtscore übertragen. Dieser Gesamtsummenscore ermöglicht Prädiktionen zum Vorliegen eines (klinisch signifikanten) PCa. Zusätzlich wird die Prostata zur möglichst genauen Angabe der Lokalisation einer Läsion in ein Schema mit bis zu 27 standardisierten Sektoren unterteilt [108].

Durch eine internationale Zusammenarbeit des American College of Radiology (ACR), European Society of Urology (ESUR) und AdMetech Foundation wurde 2015 das Scoring System im Hinblick auf globale Standardisierung in Bildakquise und Interpretation überarbeitet. Die PI-RADS Version 2 (PI-RADS v2) gewichtet die Sequenzen nicht mehr gleichwertig, sondern rückt für die PZ die DWI sowie die davon abgeleiteten ADC-maps als Leitsequenz in den Vordergrund. Die TZ wird primär in der T<sub>2</sub>w befundet, während die DCE-MR hauptsächlich für die weiterführende Klassifikation von in der DWI als PI-RADS 3 eingestuften Läsionen herangezogen wird. Abb. 3 veranschaulicht das PI-RADS Assessment. Auch das Sektorenmodell wurde überarbeitet und umfasst nun 39 Regionen, davon 36 für die Prostata und 2 bzw. 1 für die Samenblasen und den externen Sphinkter urethrae [91].

Tab. 2: PI-RADS Scores 1 – 5 und das Risiko für das Vorliegen eines PCa [91].

|                 |  |
|-----------------|--|
| PI-RADS Score 1 | „Very low (das Vorliegen eines klinisch signifikanten Karzinoms ist sehr unwahrscheinlich)“            |
| PI-RADS Score 2 | „Low (das Vorliegen eines klinisch signifikanten Karzinoms ist unwahrscheinlich)“                      |
| PI-RADS Score 3 | „Intermediate (das Vorliegen eines klinisch signifikanten Karzinoms ist fragwürdig (unklarer Befund))“ |
| PI-RADS Score 4 | „High (das Vorliegen eines klinisch signifikanten Karzinoms ist wahrscheinlich)“                       |
| PI-RADS Score 5 | „Very high (das Vorliegen eines klinisch signifikanten Karzinoms ist sehr wahrscheinlich)“             |

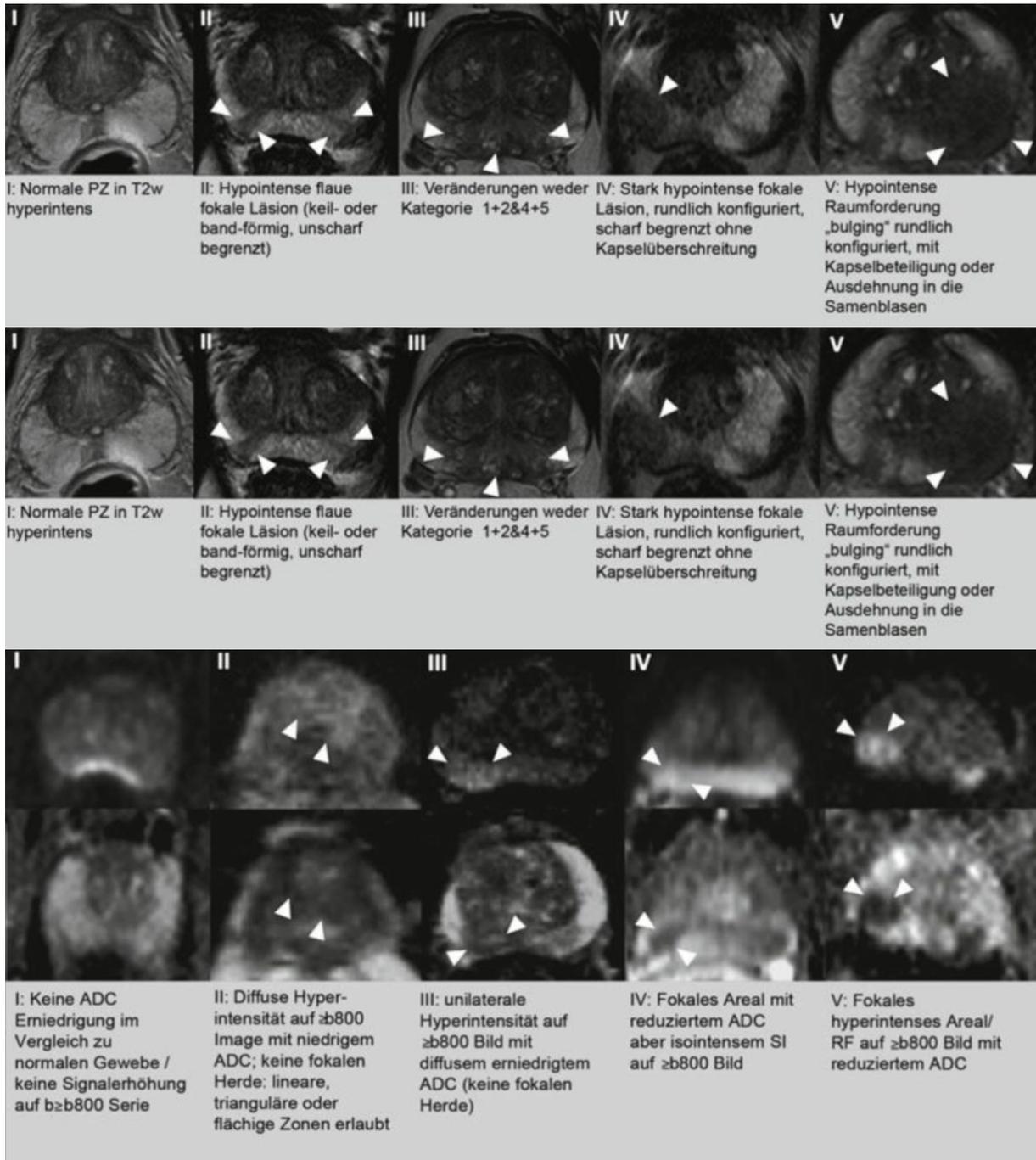


Abb. 3: Erste Reihe: PI-RADS Assessment für T<sub>2w</sub>: periphere Drüsenabschnitte. Zweite Reihe: PI-RADS Assessment für T<sub>2w</sub>: zentrale Drüsenabschnitte. Dritte Reihe: PI-RADS Assessment für DWI (ADC und hohe b-Werte) [134]; aus: Thieme Open Access.

Die Betrachtung von morphologischer hochauflösender T<sub>2w</sub> Bildgebung, DWI und DCE-MR reicht aus, um ein moderate-/high-risk PCa mit einer Diagnosegenauigkeit von 0,9 zu detektieren (dazu muss jede Läsion mit einem Gesamtscore > 3 biopsiert werden). Die Diagnosegenauigkeit für klinisch signifikante PCa kann durch den PI-RADS Score im Vergleich zur herkömmlichen Diagnostik (DRU, PSA-Wert, Prostatavolumen) von 81% auf 91% erhöht werden [135].

Die PI-RADS-Richtlinien wurden 2019 auf eine Version 2.1. aktualisiert [110]. Bei diesem Update wurden z.B. technische Bedingungen nachgeschärft und Aspekte in der

Läsionsbewertung verändert, insbesondere Kriterien für Läsionen in der TZ. Zudem wurden in die aktualisierte Version Läsionen in AFMS und zentraler Zone aufgenommen [136].

## 2.8.Klinische Problemstellung

Die Diskrepanz zwischen dem hohen Lebenszeitrisko für Männer, an einem PCa zu erkranken, und dem zugleich niedrigen Mortalitätsrisiko begründet die Notwendigkeit, zwischen therapiebedürftigen Patienten und Patienten mit einer nicht lebenszeitlimitierenden Erkrankung zu differenzieren.

Doch auch bei den behandlungsbedürftigen Patienten besteht für eine differenzierte Therapieempfehlung die Notwendigkeit einer möglichst akkuraten Risikostratifizierung: Das mittel- und langfristige Outcome eines Patienten hängt entscheidend davon ab, wie akkurat prognoserelevante Faktoren wie Tumorlokalisierung, -ausdehnung, -volumen und -aggressivität prä-interventionell eingeschätzt werden. Die möglichst schnelle und genaue Charakterisierung der Erkrankung ist Grundvoraussetzung für die Auswahl der bestmöglichen Therapiestrategie. Die Therapieansätze im Management des PCa haben sich in den letzten Jahren weiterentwickelt und verbessert. Auch wird eine zunehmende Therapieindividualisierung angestrebt. Die Therapie des PCa sollte nach der Ausprägung der Erkrankung (TNM-Stadium, GS, PSA-Wert) und individuellen Prognosefaktoren (Allgemeinzustand, Alter, Komorbiditäten des Patienten) ausgewählt werden. Die Möglichkeiten reichen von abwartenden Strategien wie Active Surveillance bei geringgradig aggressiven PCa über medikamentöse (z.B. hormonablativ) Therapien bis hin zur radikalen Prostatektomie und Bestrahlung im Falle eines fortgeschrittenen Tumorstadiums [137].

Die Differenzierung zwischen klinisch signifikanten und nicht-signifikanten PCa spielt insbesondere im Zusammenhang mit dem Therapiekonzept der Active Surveillance eine wichtige Rolle, welches 2001 erstmals beschrieben wurde und seit einigen Jahren fest im Management von PCa-Patienten etabliert ist [138]. Der Wechsel in ein „aktives“ Therapieschema erfolgt hier erst im Falle eines Tumorprogresses, in der Hoffnung, die potentiell belastende Behandlung um einige Jahre aufschieben oder sogar komplett zu vermeiden zu können [139-141].

Als eine weitere zunehmend wichtige Therapiestrategie verfolgt die personalisierte oder auch Präzisionsmedizin (engl.: „targeted therapy“ oder „precision medicine“) den Ansatz, Krankheiten zugeschnitten auf die individuellen Eigenschaften eines Patienten und seiner Erkrankung zu behandeln. Präzisionsmedizin bezeichnet einen komplexen Prozess, der mit einer breiten Grundlage an individuellen Informationen wie Genanalysen, Phänotypisierung, Biomarkern, Big Data und Feedback-Loops arbeitet. Somit soll die Wirksamkeit einer Therapie maximiert und das Auftreten unerwünschter Nebenwirkungen minimiert werden [142]. Die Onkologie ist hierfür ein prädestiniertes Fachgebiet, und in keinem anderen Fach wurden bisher so viele Ansätze aus der Präzisionsmedizin untersucht und implementiert wie in der Onkologie [143].

So beinhaltet die Präzisionsmedizin auch bei der Therapie des PCa zielgerichtete, individualisierte Behandlungsstrategien, darunter neue Genom-Technologien wie Gen-Fusionen, nicht-kodierende RNA-Biomarker, Liquid Tumor Profiling, und Genom-Editing-Technologien [137].

Um diese Informationen in der breiten Anwendung liefern zu können, werden neue diagnostische Methoden benötigt.

Ein Problem bei der präzisen Charakterisierung der Tumorerkrankung und der Auswahl von geeigneten zielgerichteten Therapien stellen klinisch bedeutsame Genmutationen dar, die auch beim PCa eine substanzielle Rolle spielen. Die intratumorale Heterogenität des PCa ist noch

nicht ausreichend verstanden [144]. In ihr liegt allerdings eine große Hoffnung auf neue Therapieansätze, da das PCa sowohl auf klinischer als auch auf zellulärer Ebene eine sehr heterogene Erkrankung ist, weshalb die bisherigen Therapieerfolge oftmals unbefriedigend sind [137].

Neue Technologien wie das Genom-Profilung haben zwar dazu verholfen, Tumoren auf genetischer Ebene besser zu verstehen und spezifische Behandlungsmethoden zu identifizieren [145]. Diagnostisch sind Aussagen über vorliegende Genmutationen und die ITH jedoch bis dato nur durch die Biopsie oder eine Resektion möglich.

Das PCa ist im organbeschränkten Stadium (Tumorstadium  $\leq$  T2c) am besten kurativ behandelbar [146]. Jedoch ist die Differenzierung zwischen einem pT2 oder einem pT3-Stadium oft schwierig, ebenso wie das pathologische Assessment des Prostatektomiepräparats oft problematisch sein kann [147, 148].

Auch andere oft herangezogene Prognosefaktoren wie Gleason-Score, PSA-Wert und TNM-Stadium gewährleisten nicht die präzise Zuordnung der Patienten in Risikogruppen [19, 149].

Die Unzulänglichkeiten der bisherigen Methoden zur Risikostratifizierung können eine fundierte Therapieentscheidung erschweren: Dem Undergrading bei vermeintlichen low-risk Patienten stehen Überdiagnostik und Übertherapie von klinisch nicht signifikanten Karzinomen gegenüber. Beide Situationen sind im klinischen Alltag relevant.

Die Patienten mit low-risk PCa von denjenigen mit klinisch signifikanten PCa zu unterscheiden, ist bereits seit langem Ziel vieler Autoren.

Im Folgenden wird dargestellt, worin die Limitationen der Untersuchungsmethoden für PCa liegen und inwiefern sie ein Risiko für Undergrading bzw. Überdiagnose bedeuten können.

### 2.8.1. Undergrading des PCa

Im Falle eines Undergradings wird das Krankheitsstadium unterschätzt. Dies bringt das Risiko für eine fortgeschrittene, ggf. metastasierte Situation mit sich, die deutlich schlechter oder nur noch durch hoch invasive Behandlungen therapierbar sein kann und die daher unbedingt vermieden werden sollte.

PSA-Wert, DRU und TRUS sind zwar geeignete Methoden zur Diagnosestellung eines PCa, jedoch wird durch sie das Tumorstadium häufig unterschätzt. Laut einer Studie steigt die Wahrscheinlichkeit für ein präoperatives Undergrading mit steigender Höhe des PSA-Serumwertes: im Bereich eines PSA-Werts von 4 – 10 ng/ml kam es bei 32% der Patienten zu einem Undergrading, im Bereich von  $>$  10 ng/ml sogar bei 47% [150].

Zudem ist bekannt, dass eine weiterführende Diagnostik bei Patienten mit einem PSA-Wert unterhalb des Cut-Offs von 4 ng/ml nicht selten doch high-grade PCa detektiert. So konnten Studien zeigen, dass die Detektionsrate von klinisch signifikanten PCa bei PSA-Werten von 2,6 ng/ml bis 4 ng/ml dieselbe war wie bei Patienten mit PSA-Werten  $>$  4 ng/ml [151, 152]. Laut einer Studie von Ohi et al. wurde bei einem Anteil von 21,7% der Patienten mit PSA-Werten zwischen 2ng/ml und 4 ng/ml ein PCa detektiert, wovon ein Anteil von rund 33% bereits ein kapselüberschreitendes PCa, und knapp 27% bereits einen Gleason Score von 4 oder 5 aufwies [153].

Auch sind PSA, DRU und TRUS zur Beurteilung lokal fortgeschrittener Karzinome und zum Staging nicht akkurat genug. Beispielsweise ist die Kapselinfiltration ein bedeutsamer prognoserelevanter Faktor. Jedoch werden bis zu 30% der Patienten mit extrakapsulärer Ausbreitung mit diesen Methoden nicht erfasst [154, 155].

Auch in einer Studie von Noguchi et al. an 222 Männern wurde der anhand von Biopsieergebnissen eingeschätzte Tumorgrad in fast jedem zweiten Fall (46%) unterschätzt [156].

Der zur Einteilung in Risikogruppen und zur Einschätzung des Rezidivrisikos verwendete Risikoscore nach D'Amico ermöglicht ebenfalls nur eine grobe quantitative Beurteilung des Rezidivrisikos mit Upgrading in bis zu 43% der Fälle [157].

Auch die Verwendung des Gleason-Scores birgt die Gefahr des Undergradings. Ein Gleason-Score  $< 6$  ist meist ein zentraler Gegenstand in den Kriterien zur Klassifikation von low-risk Tumoren. Zur Identifizierung der für eine Überwachungsstrategie geeigneten Patienten werden häufig die Epstein Kriterien herangezogen (Tumorstadium cT1c, PSAD  $< 0,15$  ng/ml, kein Gleason-Grad-4- oder -5-Muster, weniger als 3 von 6 Biopsien positiv,  $< 50\%$  Tumor in einer der Biopsien). Jedoch zeigte Epstein selbst in seiner Pionierarbeit, dass unter Anwendung seiner Kriterien nur in 73% der Fälle ein klinisch nicht signifikantes PCa korrekt diagnostiziert wurde [158].

In einer Studie von Bastian et al. wurden 237 Patienten prostatektomiert, die gemäß der Epstein Kriterien ein klinisch nicht signifikantes PCa hatten; die Diagnose konnte in 91,6% der Fälle bestätigt werden, aber bei 8,4% lag bereits ein organüberschreitendes Karzinom vor. Die Klassifikation nach Epstein Kriterien kann also für einen Anteil von 8,4% bis zu 27% der Patienten eine falsche Therapie nach sich ziehen [158, 159].

Der TRUS-Biopsie entgehen zwischen 20 % und einem Drittel der PCa [160, 161]. Im Falle eines positiven Befundes nach einer TRUS-Biopsie ist zwar der klinische Verdacht eines PCa bestätigt, aber es liegen nur unzureichende Informationen über die räumliche Ausbreitung des Tumors vor. Auch können etwaige high-grade-Tumorkomponenten der systematischen Biopsie entgehen, da das PCa ein sehr heterogener Tumor ist, der häufig in kleinen, multifokalen Herden vorliegt, was zu einem zu niedrig ausfallenden Gleason-Grad oder hoher Falsch-negativ-Rate führen kann. Die Anzahl der Foki eines multizentrischen PCa ist ein wichtiger prognostischer Faktor [144, 162].

Da zum Beispiel bei der Stanzbiopsie aus verschiedenen Regionen Proben in Form von Gewebezylindern entnommen werden, entgehen dieser Diagnostikmethode sämtliche Informationen aus anderen, nicht biopsierten Bereichen. Dadurch ist der Informationsverlust gegebenenfalls bedeutsam und Re-Biopsien unter Inkaufnahme ihrer Nebenwirkungen können nötig werden [36, 163].

Im Rahmen einer Studie an 241 Patienten mit T1c PCa, die über mindestens drei Jahre jährlich rebiopsiert wurden, kam es bei insgesamt 20% zu einer Progression des Gleason-Scores, 50% davon innerhalb der ersten zwei Jahre nach Diagnosestellung [164]. In einer weiteren retrospektiven Review-Studie an 104 Patienten kam es durch unmittelbar wiederholte Biopsie bei 27% zu einem Upgrading oder Upstaging [165]. In einer Studie an 843 Patienten kam es bei 50,2% zu einem Upgrading der TRUS-Biopsie-Ergebnisse nach der pathologischen Untersuchung der Prostatektomiepräparate. Demnach erhielt jeder zweite Patient eine Therapie basierend auf einem inkorrekt eingestuften Tumorgrad [149].

Die Daten weisen darauf hin, dass es nicht etwa kurzfristig bzw. innerhalb von drei Jahren zu einer Tumorprogression gekommen war, sondern dass die high-grade Tumorkomponenten in der bzw. den vorausgegangenen Gewebeproben schlichtweg nicht enthalten und somit auf einen Stichprobenfehler zurückzuführen waren.

Auch autolytisch zerfällt das Gewebe in einem Prostatektomiepräparat vergleichsweise schnell, und das PCa ist mit bloßem Auge kaum zu erkennen. Zum anderen ist bekannt, dass bei der histopathologischen Untersuchung eines Prostatektomiepräparats nur ein kleiner Teil des Gewebes im Resektat untersucht wird, was ebenfalls einen hohen Informationsverlust bedeuten kann [147, 166, 167]. Zusätzlich kann das Heranziehen der Prostatektomiepräparate zur Bestätigung von PCa zu einer Auswahl einer aggressiveren Läsions-Population führen, was die Prätestwahrscheinlichkeit erhöht [162].

Des Weiteren gehören Lymphknotenmetastasen zu den für die Prognose relevantesten Faktoren. Der Lymphadenektomie als Goldstandard zur Stagingdiagnostik entgehen allerdings bis zu 41% der Metastasen, da diese außerhalb des Standardfeldes lokalisiert sind [148, 168].

Wie im Abschnitt „mpMRT der Prostata“ dargestellt, spielt die MRT unter den bildgebenden Methoden zur Diagnostik von PCa eine zentrale Rolle. Die MRT-Bildgebung und darauf gestützte zielgerichtete Stanzbiopsie haben die Detektion von klinisch signifikanten PCa bereits verbessert, jedoch wird ein substantieller Teil der Karzinome dennoch nicht erkannt, wie eine Metaanalyse zeigte [169].

Eine weitere Metaanalyse von 75 Studien und 9796 Patienten zur Diagnosegenauigkeit der MRT zeigte zwar eine hohe Spezifität (0,91), aber eine niedrige Sensitivität (0,57) der mpMRT bei der Detektion von extrakapsulärer Ausbreitung bzw. zum Staging von lokalisiertem PCa, sowie eine wesentliche Unterschätzung des Tumorumfangs [155].

Obwohl das PI-RADS-Klassifikationssystem die Diagnostik des klinisch signifikanten PCa verbessert, weist es laut einer multizentrischen retrospektiven Studie dennoch als Schwachpunkt eine bestenfalls moderate bis gute Interreader-Übereinstimmung auf [170]. Laut Rahbar et al. werden bis zu 80% der PCa mit GS 6 und bis zu 28% der PCa mit GS >7 nicht detektiert [171].

Zwar haben Unterschiede in Studiendesign, Datenakquisition, Validierungskriterien und Erfahrung der Befundenden eine erhebliche Heterogenität der Studienlage zu PI-RADS v2 zur Folge [162].

Laut einigen Studien ist die Einschätzung von klinisch signifikanten PCa durch PI-RADS nicht zufriedenstellend [172-175]. Die meisten retrospektiven Studien konnten eine Korrelation zwischen höherer PI-RADS v2-Kategorie und genereller PCa-Detektionsrate sowie klinisch signifikanten PCa und zeigen [176-179]. Eine Studie zeigte jedoch einen starken Abfall der PCa-Detektionsrate in PI-RADS v2-Kategorie 5 (86,9%) und Kategorie 4 (39,1%) [180]. Diese niedrige Detektionsrate steht im Widerspruch zur PI-RADS Definition der Kategorie 4 („Vorliegen eines klinisch signifikanten Karzinoms wahrscheinlich“) [91].

In einer weiteren Studie wurde untersucht, ob die PI-RADS-Klassifikation sich eignet, um Vorhersagen zur Multifokalität von PCa zu treffen, welche durch die mpMRT nicht erfasst wird. Bei 30% der Patienten wurden kleinere, klinisch signifikante PCa abseits einer Indexläsion nicht erkannt. Darüber hinaus stiegen mit höherem PI-RADS Score der Indexläsionen die Rate und Ausmaße der in der mpMRT nicht erfassten signifikanten PCa deutlich an. Falsch-negative Befunde korrelieren demnach in hohem Maße mit dem PI-RADS Score der Indexläsion [181].

### 2.8.2. Überdiagnose des PCa

Die Überdiagnose bezeichnet die Feststellung einer Erkrankung, die im Laufe des Lebens des Patienten nicht symptomatisch bzw. lebenszeitlimitierend geworden wäre [182].

Die Wahrscheinlichkeit, dass es bei PCa zu Überdiagnose kommt, ist vergleichsweise höher als bei anderen Tumorentitäten, da beim PCa durch die hohe Prävalenz indolenter Tumoren in oft höherem Lebensalter andere Ursachen eher todesursächlich sind [183].

Eine weitere große Studie konnte zeigen, dass bei Männern, die im Zeitraum von 10 Jahren alle 2-4 Jahre einem PSA-Screening unterzogen werden, bei mehr als 15% mindestens ein falsch-positives Ergebnis vorliegt [184]. Die Anzahl falsch-positiver Ergebnisse und die Komplikationsrate von Biopsien steigen in höherem Lebensalter [185].

Von einer Behandlung dieser Tumorerkrankungen profitieren die Männer nicht, vielmehr werden die teilweise erheblichen Nebenwirkungen überflüssigerweise riskiert („Übertherapie“) [5].

Dies ist im Falle von PCa insofern bedeutsam, als dass das lokalisierte PCa standardmäßig mit radikaler Prostatektomie und Radiotherapie behandelt wird. Diese Therapien können mit irreversiblen und die Lebensqualität erheblich beeinträchtigenden Komplikationen

einhergehen, wie z.B. Urin- und Stuhlinkontinenz, erektiler Dysfunktion, Urethrastrikturen, Harnverhalt und Harnwegsinfekten [11, 186]. Laut einer Studie haben z.B. noch 5 – 10 Jahre nach Therapie bis zu 48% der Patienten Miktionsbeschwerden, 5% - 14% haben stuhlgangsbezogene Beschwerden und 40% - 74% leiden unter sexueller Dysfunktion [187]. Entsprechend sollte unbedingt vermieden werden, Patienten mit indolenten und langsam-fortschreitenden Karzinomen überflüssigerweise einer invasiven, nebenwirkungsbehafteten Therapie zu unterziehen [188-190].

Zusätzlich ist die psychische Belastung des Wissens um eine Krebserkrankung, die das Leben andernfalls nicht beeinträchtigt hätte, bedeutsam [183].

Laut Prensner et al. weist der PSA-Test eine Spezifität von 20% - 40%, und eine Sensitivität von 70% bis 90% auf, ausgehend von einem Schwellenwert des PSA-Levels von 4 ng/ml [191]. Dies lässt sich dadurch erklären, dass das PSA kein tumor- sondern ein gewebespezifischer Marker ist, sodass sowohl maligne als auch benigne Prozesse zu erhöhten PSA-Serumkonzentrationen führen können. Dies kann falsch-positive Ergebnisse und damit zu Überdiagnose bedingen. Beispielsweise steigt der PSA-Wert bei Größenzunahme der Prostata um 3,2% pro ml Prostatavolumen. Altersabhängig steigt er bei benigner Prostatahyperplasie um 2,6% pro Lebensjahr, sodass bei Männern mit BPH entsprechend die Spezifität der Untersuchung des PSA-Werts beeinträchtigt wird [65, 192]. Einen kurzfristigen PSA-Anstieg können auch Prostatitiden sowie Manipulationen wie TRUS, DRU oder Biopsie herbeiführen [65, 193].

Die Nachweisrate eines PCa liegt im PSA-Wert-Bereich von 4 ng/ml bis 10 ng/ml nur bei etwa 25%, sodass sich Patienten gegebenenfalls wiederholten Stanzbiopsien bis hin zur Sättigungsbiopsie (Entnahme von 24 oder mehr Proben) unterziehen, ohne dass ein PCa entdeckt wird [161, 194].

In der im Abschnitt „Undergrading“ bereits angeführten Studie von Noguchi et al. an 222 Männern wurde außerdem der Tumorgrad anhand von Biopsieergebnissen bei 18% der Patienten überschätzt [156].

Darüber hinaus konnte in einer weiteren Studie bei einem Anteil von 33% der Patienten mit PI-RADS-5-Läsion kein klinisch signifikantes PCa, sondern nur gutartiges Gewebe bioptisch nachgewiesen werden, was die Limitation der Begutachtung von MRT-Aufnahmen verdeutlicht [195].

Eine Studie zu Screening, Überdiagnose und Übertherapie von PCa in England und den USA sowie den Auswirkungen für das Gesundheitssystem in Zukunft zeigte, dass 39% der Kosten für Screening und Therapie des PCa auf überdiagnostizierte PCa entfallen [14].

Eine große Studie kam zu dem Ergebnis, dass zwischen 20% und 50% der durch Screening erfassten PCa überdiagnostiziert sind [185]. Laut der Einschätzungen einer Reihe von weiteren Studien beläuft sich die Rate der Überdiagnose von klinisch nicht signifikanten PCa sogar auf bis zu 67% [14, 196-198].

Der Trade-Off von Früherkennung und Überdiagnose von indolenten PCa kann demnach zu einem erheblichen Anteil von Übertherapie mit erheblicher Einschränkung der Lebensqualität führen [199].

Zu der dargestellten Problematik kommt die Tatsache hinzu, dass in Deutschland jährlich mehr als 70 Millionen MRT- und CT-Aufnahmen generiert werden. Dieser Masse an Untersuchungen steht eine vergleichsweise kleine Facharztgruppe an Radiologinnen und Radiologen gegenüber (weniger als 10.000 im Jahr 2019), sodass eine automatische Auswertung und Unterstützung der manuellen Diagnoseerbringung zunehmend wünschenswert wird [200, 201].

## 2.9. Radiomics

Computergestützte Diagnosesysteme (CDS) sollen idealerweise Prozesse im klinischen Alltag automatisieren, beschleunigen, präzisieren und verlässlicher machen, indem mithilfe geeigneter Verfahren komplexe Zusammenhänge erkannt und Schlussfolgerungen abgeleitet werden. Die Effizienz und Geschwindigkeit Computer-gestützter Analysen können dabei die Fähigkeiten des Menschen weit übersteigen [202, 203].

Hinzu kommt das wachsende Potential neuer Computertechnologien, das die Entwicklung automatisierter, quantitativer Bildanalysemethoden unter Anwendung von künstlicher Intelligenz (KI) und maschinellem Lernen ermöglicht.

Ein wichtiges Konzept aus diesem Bereich in der Radiologie wird als „Radiomics“ bezeichnet.

### 2.9.1. Radiomics: Begriff und Definition

Der Begriff „Radiomics“ wurde von Aerts et al. geprägt [204] und ist ein Neologismus, der sich aus den Worten „radiology“ und „omics“ zusammensetzt. Das Wort „Omics“ ist ebenfalls ein Neologismus und stammt aus dem Forschungsbereich der Biologie. Er dient als Sammelbegriff für molekularbiologische Hochdurchsatzmethoden, in denen komplexe biologische Systeme und ihre Teilkomponenten untersucht werden, z.B. aus den Bereichen der Genomics, Proteomics, Metabolomics, Secretomics, Pharmacogenomics, Transcriptomics etc. [205].

Radiomics ist eine neue Methode der automatisierten Bildanalyse. Hierbei werden bildmorphologische Merkmale aus radiologischen Bildern extrahiert, quantifiziert, selektiert, und für Klassifikationsmodelle verwendet. Dem liegt die Annahme zugrunde, dass verschiedene Zelltypen, Stoffwechsel- oder molekularbiologische Prozesse eines Gewebes durch digitale Bildgebung in Daten verschlüsselt werden. Diese Bilddaten können zum einen visuell analysiert und beurteilt werden, jedoch sind der subjektiven Bildbetrachtung nicht alle Informationen, die in einem Bild kodiert sind, zugänglich. Im Rahmen einer Radiomics-Analyse werden aus digitalen Bilddaten eine Vielzahl von Bildmerkmalen in Form von quantitativen, hochdimensionalen Daten extrahiert. Die Auswertung solcher Bildinformationen hat einen Erkenntnisgewinn und ein tieferes Verständnis von Zusammenhängen zwischen Bereichen wie der Molekulargenetik, Pathophysiologie, Mikroanatomie und Radiologie zum Ziel, die sich z.B. die personalisierte Medizin zunutze machen könnte [206, 207].

Eine Arbeit von Segal et al. legte im Jahr 2007 den Grundstein für die Radiomics-Entwicklung. Die Autoren zeigten in ihrer Studie, dass Bildmerkmale aus einer CT-Bildgebung, die sie als „imaging traits“ und „radiographic features“ bezeichneten, mit Genexpressionsmustern von hepatozellulären Karzinomen korrelieren. Indem die errechneten Merkmale kombiniert wurden, konnten Rückschlüsse über die Zellproliferation und somit über die Prognose der Patienten gezogen werden [208].

Weitere Publikationen zu Radiomics wurden 2012 veröffentlicht. In einer Grundstein-legenden Arbeit auf dem Gebiet postulierten Lambin et al., dass die quantitative Analyse von Bildmerkmalen mehr und effizienter Informationen gewinnt als die alleinige visuelle Befundung der Bilder durch eine Ärztin bzw. einen Arzt. Dies gelte auch für Prognosen, die sich aufgrund einer Korrelation von nicht-invasiv gewonnenen phänotypischen Informationen mit genotypischen und klinischen Parametern treffen ließen [209].

Zwei Jahre später lag die bereits erwähnte, wegweisende Arbeit von Aerts et al. vor, in der 440 Radiomics Features von 1019 Patienten mit Bronchialkarzinom- oder Kopf-Hals-Tumoren mit Daten zu genetischem Profil, TNM-Klassifikation und Gesamtüberleben verglichen wurden. Es stellte sich eine sogenannte Radiomics-Signatur, also eine bestimmte integrative

Kombination verschiedener Merkmale, mit vier Features heraus, welche der prognostischen Aussagekraft der TNM-Klassifikation und dem Tumorzellen beim Bronchialkarzinom und bei Kopf-Hals-Tumoren gleichkam bzw. überlegen war [204]. Unter Nutzung derselben 440 Features konnte die Gruppe um Aerts 2017 in einer Studie an Lungentumoren den KRAS-Mutationsstatus und den EGFR-Status anhand von CT-Aufnahmen vorhersagen, und damit essentielle Informationen für die Therapieauswahl liefern [210].

Laut einer Review-Studie wurden zwischen 2017 und 2018 in der PubMed/MEDLINE Database mehr als 500 Publikationen zum Thema Radiomics veröffentlicht, wobei einige Radiomics-Ansätze für eine Vielzahl von Anwendungsbereichen, darunter auch die Prostatakrebsdiagnostik, äußerst vielversprechend erscheinen [211].

### 2.9.2. Funktionsprinzip von Radiomics

Die Befundung von bildgebenden Verfahren als Bilder fokussiert sich herkömmlicherweise auf visuelle, qualitative Kriterien, wie beispielsweise Abgrenzbarkeit, Form, Größe und Kalzifizierung einer Läsion [202]. Doch bei dieser Vorgehensweise bleiben die latenten biomedizinischen Bilddaten unberücksichtigt, von denen digitale Datensätze der bildgebenden Untersuchungen jedoch eine große Anzahl beinhalten [212, 213].

In radiologischen Bilddaten ergibt sich die räumliche Bildauflösung aus der Anzahl der Bildpunkte und der Größe des erfassten Bereiches.

Jede radiologische Schnittbildaufnahme enthält eine große Zahl an Bildpunkten pro Schicht, z.B. 256 x 256 oder 512 x 512. Jeder Bildpunkt repräsentiert ein Volumenelement, dem ein Zahlenwert zugewiesen ist. Die Zahlenwerte werden wiederum zur Betrachtung in Graustufen dargestellt. Während ein Bildpunkt aufgrund einer typischen Speichertiefe von 12 bit prinzipiell 4096 Abstufungen aufweisen kann, ist das visuelle System des Menschen hingegen nur zur Unterscheidung von bis zu 30 Graustufen in der Lage [214].

Dies stellt eine Limitation in der Bildbefundung dar, die z.B. in der CT durch verschiedene Fensterungen (z.B. Lungenfenster, Knochenfenster) versucht wird zu kompensieren.

Auch ist bekannt, dass die Relation benachbarter Farben oder Grauwerte die Wahrnehmung des menschlichen Auges täuschen kann: Nebeneinander liegende Grauwertflächen werden in Wechselwirkung zueinander wahrgenommen, was zu Herauf- oder Herabsetzen des Kontrastes und damit Fehleinschätzungen durch das menschliche Gehirn führt. Als Beispiel für diese Problematik lässt sich eine über einen längeren Zeitraum beobachtete Läsion anführen, deren Dichte sich z.B. durch eine Therapie verändert. Ändert sich die Helligkeit der umliegenden Gewebe gleichermaßen, kann die Läsion als unverändert beurteilt werden [207].

Radiomics-Analysen setzen unter anderem genau an dieser Stelle an: Das bloße Auge der/der befundenden Ärztin bzw. Arztes wird durch eine computerisierte Hochdurchsatz-Analyse ergänzt.

Unter dem Begriff „Bildmerkmale“ im Zusammenhang mit Radiomics wird also nicht das von menschlichem Auge und Gehirn gesehene visuelle Bild verstanden, sondern Bilddaten, die dem menschlichen Auge möglicherweise verborgen bleiben. Dies können z.B. quantitative Formmerkmale, komplexe mehrdimensionale Muster, feinste Graustufen, Homogenität, Heterogenität, interpixel-Relationen, Texturen, spektrale Eigenschaften usw. in einer Region of Interest (ROI) sein. Die Bestimmung des Mittelwerts der Signalwerte innerhalb einer ROI sind ein Beispiel für eine einfache statistische Auswertung, die bereits fest in den klinischen Workflow integriert ist.

Durch Identifikation und Analyse oben aufgeführter Bildmerkmale werden sogenannte radiomischen Features festgelegt, die diese Merkmale somit beurteilbar machen. Mithilfe von maschinellem Lernen können diese anschließend in der Auswertung mit Zielparametern

verknüpft werden. So kann beispielsweise die Ausprägung der intratumoralen genetischen Heterogenität, verschlüsselt in eine Radiomics-Signatur, in Zusammenhang mit Therapieansprechen oder Prognose gebracht werden.

Ziel ist es also, aus diesen Features klinisch relevante Informationen mit differentialdiagnostischem oder prognostischem Wert abzuleiten, bzw. Bildmerkmale zu ermitteln, die sich z.B. als Indikatoren für pathologische Prozesse verwenden lassen. Dies basiert auf der Annahme, dass sich pathophysiologische und phänotypische Tumoreigenschaften mit Bildgebungstechniken erfassen und korrelieren lassen. Bestimmte Features könnten demnach spezifisch für bestimmte Pathologien sein. Somit lassen sich diese Merkmale auch als „bildgebende Biomarker“ oder „bildgebender Phänotyp“ verstehen [209, 215].

Ein Zusammenhang zwischen quantitativen Bildmerkmalen und klinischen Parametern wie der Präsenz eines Tumors oder Behandlungsergebnis konnte bereits gezeigt werden: Radiomics Features lassen sich beispielsweise mit Genexpression, Hormonstatus, Genmutationen, histopathologischem Grading, Gefäßdichte, Tumormetabolismus und Aggressivität korrelieren [206, 216].

### 2.9.3. Radiomics Feature-Klassen

Radiomics Features lassen sich verschiedenen Klassen zuordnen.

In der Radiologie werden z.B. qualitative semantische Features zur Beschreibung von Läsionen verwendet [217]. Hingegen werden quantitative Features mithilfe geeigneter Software von mathematischen bzw. informatischen Algorithmen extrahiert. Solche Features können unterschiedliche Komplexitätsgrade aufweisen [218]. Sie lassen sich in First-Order Statistics, Second-Order Statistics und Higher-Order Statistics einteilen.

First-Order Statistics beschreiben neben Größen wie Maximum, Minimum, Median, oder Mittelwerte der Voxel-Intensitäten in einem Bild auch die Asymmetrie des Intensitäts-Histogramms, Uniformität oder Kurtosis (Wölbung). Die räumliche Beziehung der Voxel-Werte wird dabei unberücksichtigt gelassen.

Als Second-Order-Statistics gelten Features, die unter Berücksichtigung räumlicher Aspekte im Bild errechnet werden. Sie schließen dabei die Texturmerkmale mit ein. Sie können z.B. die Heterogenität bzw. Homogenität einer ROI charakterisieren.

Higher-Order-Statistics werden anhand von Daten berechnet, die von den Primärbildern abgeleitet wurden, oder nach Anwendung von statistischen Methoden oder Filtern, beispielsweise der Wavelet Transformation oder des Laplace-Filters zur Kantendetektion. Ziel ist es dabei, z.B. Muster herauszuarbeiten, Rauschen zu reduzieren, oder bestimmte Strukturen hervorzuheben. So erhält man Features, die sich zur weiteren Verarbeitung in Machine Learning (ML; Erläuterung im folgenden Absatz) Workflows eignen [207, 219].

Quantitative Features können des Weiteren in Untergruppen kategorisiert werden:

- **Shape Features:**  
Shape Features beschreiben die Form einer ROI und ihre geometrischen Eigenschaften in 2D oder 3D, zu denen Durchmesser, Volumen, Oberfläche, Spikularisierung oder Sphärizität zählen. Diese Features sind unabhängig von Graustufenverteilungen innerhalb der ROI. Beispielsweise weist das Oberflächen/Volumen-Verhältnis eines spikulierten Tumors höhere Werte auf als das eines runden Tumors desselben Volumens. Somit lassen sich über Shape-Features Gewebe (z.B. ein Tumor) charakterisieren.

- Textur-Features:

Die Quantifizierung anhand von Bildeigenschaften mit Texturmerkmalen wurde bereits in den Achtzigerjahren angewendet. Generell liegt der Texturanalyse meist die Charakterisierung der räumlichen Verteilung von (Intensitäts-) Werten zugrunde. Gängige Textur-Features sind die sogenannten Haralick-Features, die davon ausgehen, dass die Textur eines Bildes durch die räumlichen Beziehungen der unterschiedlichen Grauwerte charakterisierbar ist [220]. Sie quantifizieren beispielsweise die Häufigkeit von Voxel mit derselben Intensität oder einem bestimmten Intensitätsunterschied in einer festgelegten Strecke und Richtung, und berücksichtigen so die räumliche Beziehung jedes Bildpunktpaars im Bild. Diese Beziehungen werden durch Grauwertematrizen (Gray-Level Co-Occurrence Matrix, „GLCM“) beschrieben, aus denen sich wiederum weitere Texturmerkmale ableiten lassen. Die Grauwertematrix ist eine wichtige Technik in der digitalen Bildverarbeitung [219, 220].

Als ein Beispiel für Second-Order Statistics quantifiziert die Gray-Level Difference Matrix (GLDM) die absolute Differenz zwischen Paaren von Pixelwerten. Anhand der GLDM lassen sich weitere Deskriptoren wie Kontrast, Mittelwert, Varianz, Entropie, oder Cluster-Schattierung berechnen.

Die Gray-Level Size Zone Matrix (GLSZM) beispielsweise quantifiziert Graustufenzonen in einem Bild. Diese Zonen werden durch die Anzahl untereinander verbundener Voxel mit der gleichen Graustufen-Intensität definiert. Im Gegensatz zu GLCM ist die GLSZM unabhängig von der Ausrichtung der Voxels, und es wird nur eine Matrix für alle Richtungen in der ROI berechnet [221, 222].

Die Anwendung von Filtern kann bei der Texturanalyse bzw. Textur-Feature-Extraktion zum Einsatz kommen, um die Hyperintensität oder Heterogenität in einer ROI zu charakterisieren und damit Rückschlüsse auf die abgebildeten Gewebe zuzulassen.

Auch Histogramm-basierte Features (z.B. Histogram of oriented gradients (HOG) oder Local binary pattern (LBP) wurden bereits erfolgreich in Texturanalysen angewendet [223, 224].

Die Texturanalyse kann insbesondere bei MRT Informationen über strukturelle Gegebenheiten eines abgebildeten Gewebes erbringen, die vom menschlichen Auge nicht erfasst werden können [225].

## 2.10. Künstliche Intelligenz: Grundprinzipien von Machine Learning

*„I predict that within 10 years no medical imaging study will be reviewed by a radiologist until it has been pre-analyzed by a machine.“ - R. Nick Bryan, 2016 RSNA News 26:4-6*

Inzwischen kommt in diesem Zusammenhang mehr und mehr auch die künstliche Intelligenz zur Anwendung. Sie beschreibt Anwendungen, bei denen Maschinen Leistungen erbringen, die menschlichen Intelligenzleistungen entsprechen [226].

Sie bietet die Möglichkeit, nicht mehr nur vom Menschen vorab definierte Features zu extrahieren: Mithilfe von KI können Algorithmen während ihres Lernprozesses selbst erlernen, welche Features für eine bestimmte Fragestellung relevant sein können, und ihre Analyse darauf ausrichten. Das „Machine Learning“ (dt.: maschinelles Lernen) gilt als Schlüsseltechnologie der KI [227].

Aktuell beschäftigt sich die Wissenschaft vor allem damit, die Sicherheit, Robustheit (also wie effektiv ein Algorithmus auf unabhängige, aber ähnliche Daten übertragbar ist), und Transparenz zu optimieren, sowie den Trainingsaufwand zu reduzieren.

Im Bereich der „diagnostischen künstlichen Intelligenz“ stehen eine verbesserte und treffsichere Detektion von Pathologien, die Beantwortung eindeutiger Fragestellungen und das Stellen von (Differential-) Diagnosen im Vordergrund. In der „prädiktiven künstlichen Intelligenz“ geht es darum, zuverlässige, objektivierbare Vorhersagen über den Krankheitsverlauf von Patienten treffen zu können, z.B. wie gut eine Tumorerkrankung auf eine bestimmte Therapie ansprechen wird [228].

Machine Learning beschreibt die künstliche Generierung von Wissen aus Erfahrung. Ein System lernt anhand von Beispielaufgaben, verbessert sich und kann sein Wissen anschließend verallgemeinern: aus dem erworbenen Wissen wird ein Modell entwickelt, das anschließend auf unbekannte Daten gleicher Art angewendet werden kann.

Der Kern der klassischen ML-Ansätze besteht aus Mustererkennung durch Auswahl geeigneter Features, die Charakteristika in Bezug auf eine bestimmte Fragestellung am besten wiedergeben, sowie der Auswahl eines für die gewünschte Klassifizierung geeigneten Algorithmus.

Die Güte einer Klassifizierung ist abhängig von der differenzierten Auswahl der für die Fragestellung möglichst spezifischen Features, und nicht etwa von der Berücksichtigung einer möglichst großen Anzahl von Features [229]. Im Gegenteil kann eine zu große Anzahl an berücksichtigten Features zum Problem des sogenannten Overfittings (Überanpassung) führen. Dabei erlernt das Modell anhand der zu großen Auswahl an Parametern komplexe, aber unnötige Regeln, die die Generalisierbarkeit des Ergebnisses durch Überanpassung an die Trainingsdaten einschränken bzw. dieses unbrauchbar machen.

Eine anschließende Validierung des Klassifikationsmodells an einem Testdatensatz gibt Aufschluss darüber, ob das Modell generalisierbar ist [230].

### 2.10.1. Klassifikationsmodelle

Im Folgenden werden einige geläufige Klassifikationsverfahren näher erläutert.

- Classification and Regression Tree (CART) und Random Forest-Klassifikator

Der Classification and Regression Tree („CART“) ist ein wichtiger prädiktiver Algorithmustyp des Machine Learnings, der bereits 1984 erstmals publiziert wurde [231]. Er bildet beispielsweise die Grundlage für den Random Forest-Klassifikator (RF), der eine vielgenutzte Technik im Machine Learning ist.

Der CART ist ein Entscheidungsbaum zur Lösung von binären Problemen. Er zeichnet sich dadurch aus, dass er nur Regeln zur Durchführung bzw. Lösung von genau einer Klassifikation bzw. Fragestellung enthält. Dabei ist er als Sequenz von Fragestellungen strukturiert, bei der die Antwort auf jede Frage die Grundlage für die Beantwortung der nächsten Frage darstellt. Bei der graphischen Darstellung entsteht durch die Verzweigungen ein Baumdiagramm.

Das Ergebnis bzw. der Output eines solchen Klassifikators kann sowohl eine Klassifizierung als auch eine Regression sein [230, 232].

Um die Klassifikationsgüte zu erhöhen, lassen sich CART kombinieren. Somit entsteht ein Entscheidungswald („decision forest“). Während ein einzelner Entscheidungsbaum in keiner optimalen Prädiktion resultieren muss, ist dies bei einer Mehrheitsentscheidung mehrerer Bäume sehr wohl möglich.

Ein Entscheidungswald-Klassifikator besteht aus multiplen Entscheidungsbäumen, wobei jeder dieser Bäume auf einem Trainings-Datensatz aufbaut und parallel verarbeitet werden kann. Diese Entscheidungsbäume werden unabhängig voneinander entwickelt und dienen der Darstellung von Entscheidungsregeln. Schlussendlich werden die Prädiktionen aller Entscheidungsbäume kombiniert und ein Durchschnitt gebildet. Die „Krone“ jedes dieser Bäume besteht aus Knotenpunkten, und jeder Knotenpunkt teilt sich an einer Teilungsstelle, die anhand aller Features oder eines Subsets von Features aus den Trainingsdaten festgelegt wird [233]. Indem z.B. ein bestimmter Baum auf ein Subset von Features begrenzt wird, wird eine Zufälligkeit („Randomness“) generiert („Random Forest-Klassifikator“). Dadurch können Varianz und Overfitting reduziert und Generalisierbarkeit und Robustheit verbessert werden [230]. Die Kombination von Entscheidungsbäumen mit neuronalen Netzwerken ermöglicht darüber hinaus eine noch höhere Klassifikationsgüte.

Abb. 4 veranschaulicht einen solchen Random Forest-Klassifikator.

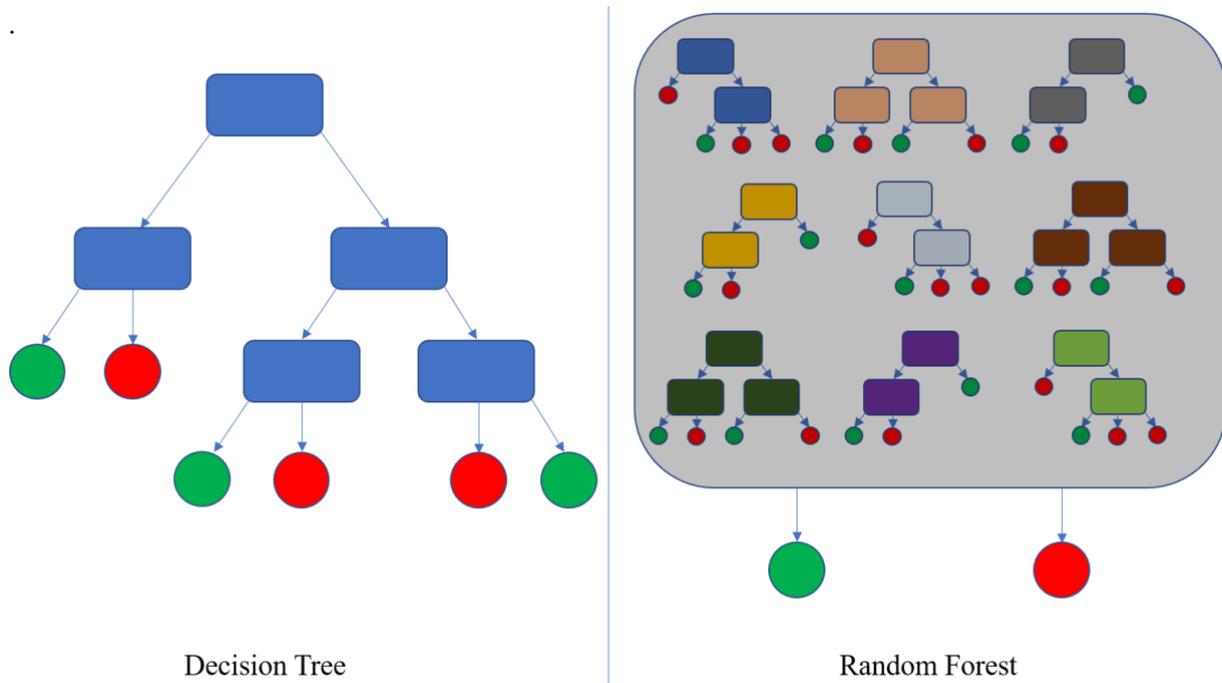


Abb. 4: Funktionsweise eines Random Forest-Klassifikators mit mehreren Entscheidungsbäumen.  
 Licensed under: *Jeremybeauchamp*, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0>>, via  
 Wikimedia Commons. Heruntergeladen von:  
[https://upload.wikimedia.org/wikipedia/commons/d/d8/Decision\\_Tree\\_vs.\\_Random\\_Forest.png](https://upload.wikimedia.org/wikipedia/commons/d/d8/Decision_Tree_vs._Random_Forest.png);

- Logistische Regression

Die logistische Regression (LR) stellt als klassische statistische Methode einen Spezialfall der Regressionsanalyse dar und wird angewendet, um ein kategoriales, nominalskaliertes Kriterium vorherzusagen. Ihre Anwendung ist also insbesondere dann sinnvoll, wenn die abhängige Variable nur wenige und gleichgestellte Ausprägungen besitzt.

Die LR ist ein äußerst robustes und vielseitig anwendbares Klassifikationsverfahren. Sie testet, ob zwischen einer abhängigen und mehreren unabhängigen Variablen ein Zusammenhang besteht. Der spezifische Einfluss von jeder unabhängigen Variablen kann mit LR quantifiziert werden [234].

Bei der Selektion der unabhängigen Variablen können in der Medizin z.B. Faktoren wie klinische Erfahrungswerte, statistische Analysen, empirische Untersuchungen usw. miteinfließen. Wichtige Schritte bei der Anwendung von LR sind die Auswahl der unabhängigen Variablen bezogen auf das angenommene Ergebnis und die Auswahl einer geeigneten Strategie zur Entwicklung eines Modells.

Auch beim LR-Modell muss die Komplexität des Modells begrenzt werden (z.B. durch Begrenzung der Anzahl der Variablen), um ein Overfitting zu vermeiden.

Vor dem Ableiten von Schlussfolgerungen aus den Analyseergebnissen muss die interne (d.h. Reproduzierbarkeit im selben Datensatz) und externe Validität (d.h. Generalisierbarkeit auf neue, unbekannte Datensätze) des Modells geprüft werden [234, 235].

- Support Vector Machine

Die Stützvektormaschine (Support Vector Machine, SVM) ist eine weitere mathematische Methode des Machine Learnings, deren typisches Anwendungsgebiet die Mustererkennung ist. Ihre Entwicklung in den 1990er Jahren stellte einen Durchbruch in der Weiterentwicklung des maschinellen Lernens dar [227]. Sie kann als Algorithmus in vielfältigen Anwendungsgebieten Objekte analysieren und mit großer Treffsicherheit Klassen zuordnen, wie z.B. in der Bilderkennung. Dabei werden die Verfahren der linearen und nicht-linearen Objektklassifizierung verwendet, die von Vektoren in einem Vektorraum repräsentiert werden. Die SVM stützt sich auf eine gute theoretische Basis, und sie gilt insbesondere bei der Verarbeitung hochdimensionaler Daten als genauer als andere Klassifikationssysteme. In der Medizin wird die SVM bereits zur Subtypisierung und Klassifizierung von Tumorgenomen verwendet [236]. Ebenso wurde die SVM bereits erfolgreich zur Klassifizierung des malignen Melanoms verwendet [237].

Zudem wurde die SVM im Kontext des PCa bereits erfolgreich eingesetzt: Z.B. bei der schwierigen Charakterisierung benigner Prostatahyperplasie und der Detektion von kleinen PCa in der TZ konnten Wu et al. einen ML-Ansatz (SVM-Klassifikator und lineare Regression als Modelle) entwickeln, der eine hohe Diagnosegenauigkeit aufwies [238]. Auch in einer weiteren Studie erwies sich der SVM-Klassifikator als am besten [239].

- Boosted-Tree-Algorithmus

Der Boosted-Tree-Algorithmus „XGBoost“ ist ein auf Entscheidungsbäumen basierender Algorithmus mit Gradientenverstärkung, mit dem sich Zielvariablen bestimmen und Schätzungen treffen lassen [240, 241]. Das Konzept wurde im Jahr 1990 geprägt [242]. Der Baumalgorithmus fügt so lange weitere Modelle hinzu, bis keine weitere Verbesserung der Prädiktionen mehr eintritt. Mit jedem hinzugefügten Modell wird versucht, die noch existierenden Fehler zu modellieren und damit im nächsten Schritt zu minimieren. Die endgültige Vorhersage entsteht aus der Verknüpfung dieser Bäume.

Somit werden zuverlässig auch komplexe Aufgaben des Machine Learnings der Bereiche Rangfolge, Regression und Klassifizierung gelöst [243, 244].

Die auch in dieser Studie verwendete Programm „XGBoost“ für überwachtes Machine Learning konnte in verschiedenen medizinischen Anwendungen State-of-the-art Ergebnisse erbringen und wurde bereits in mehreren Machine Learning-Wettbewerben ausgezeichnet [241, 245-247].

## 2.11. Methodologische Grundlagen zur Gewinnung von Radiomics Features und zur Entwicklung eines Machine Learning Modells

Ein Computeralgorithmus trainiert anhand von Trainingsbeispielen, Muster zu erkennen und Schlussfolgerungen abzuleiten. Mit der aus den Beispielen gewonnenen Erfahrung können anschließend auch unbekannte Daten beurteilt werden. Durch das Formulieren eines Minimierungsproblems kann ein Fehler zwischen der Vorhersage des Algorithmus und des tatsächlichen Werts errechnet werden. Um diesen zu verkleinern, passt das Modell seine

Parameter entsprechend immer weiter an; dies beschreibt den Lernprozess eines ML-Algorithmus.

Die Trainingsbeispiele setzen sich typischerweise aus Beobachtungen als Inputs (beispielsweise MRT-Datensätze wie in dieser Studie) und diesen zuzuordnenden Zielwerten bzw. Outputs (z.B. Tumortyp, Progress einer Tumorerkrankung, Therapieansprechen usw.) zusammen. Die Beobachtungen werden bei einer Radiomics-Analyse in Merkmalsvektoren (Form einer Läsion, Größe, Texturparameter etc.) übersetzt. Die Zielwerte können kategorischer Art (z.B. „maligne oder benigne“ oder die Klassifikation eines Tumortyps) oder skalarer Art (z.B. Tumorstadium) sein.

Es können verschiedene Lernparadigmen unterschieden werden, zu denen überwachtes („supervised“), unüberwachtes („unsupervised“) und verstärkendes („reinforcement learning“) gehören, wie auch Mischformen.

Beim überwachten Lernen werden Daten-Sets bestehend aus einer Beobachtung und einem Zielwert vorgegeben. Das System sucht dann nach Merkmalen, die die Daten voneinander unterscheiden und in Gruppen differenzieren. Im Anschluss und auf die Radiologie übertragen, kann das Modell dann neuen MRT-Aufnahmen Zielwerte zuordnen, wie z.B. einer Läsion die Kategorie „maligner Tumor“. In der Medizin erreichen ML-Modelle bei spezifischen Aufgabenstellungen teilweise bereits die Diagnosegenauigkeit einer Ärztin bzw. eines Arztes. Anzuführen sind hier die erfolgreiche Detektion von Läsionen in Mammographien, die Differenzierung zwischen Naevus und malignem Melanom, die Befundung von spinalen MRT z.B. in Bezug auf Spinalkanalstenosen, Spondylolisthesen und Bandscheibenschmälerung, oder die Beurteilung diabetischer Retinopathie und kardiovaskulärem Risiko anhand von Funduskopie-Aufnahmen des Auges [248-253]. Abb. 5 zeigt einen Fall, bei dem ein Computer-assistiertes Diagnosesystem bei der Befundung einer mpMRT der Prostata einen Vorteil erbrachte.

Die Limitation liegt darin, dass hierbei bereits bekannte diagnostische Kriterien vorausgesetzt werden, und dass über diese Kriterien hinaus eine Erkennung weiterer, neuer Gruppen nicht möglich ist [254].

An diesem Punkt setzt das unüberwachte Lernen an, bei dem durch die Methode des „Clustering“ Muster in den Eingabedaten erkannt und Daten entsprechend gruppiert werden [255]. Es werden nur Eingabewerte angeboten, die nicht vorab klassifiziert bzw. Zielwerten zugeordnet wurden, sodass das System in einem ersten Schritt selbst festlegt, welche Merkmale den Daten gemeinsam sind, und anhand dieser selbstständig eine Gruppierung durchführt.

Auf die Radiologie übertragen könnte ein System also idealerweise selbst Features identifizieren, die benigne von malignen Läsionen unterscheiden, und die Läsionen selbstständig entsprechend als benigne oder maligne klassifizieren. Während also beim überwachten Lernen Modelle ausgehend von vom Menschen vorgegebenen Beispielaufgaben mit gewünschter Lösung („Ground truth“) trainiert werden, entfällt dieser Schritt beim unüberwachten Lernen [256].

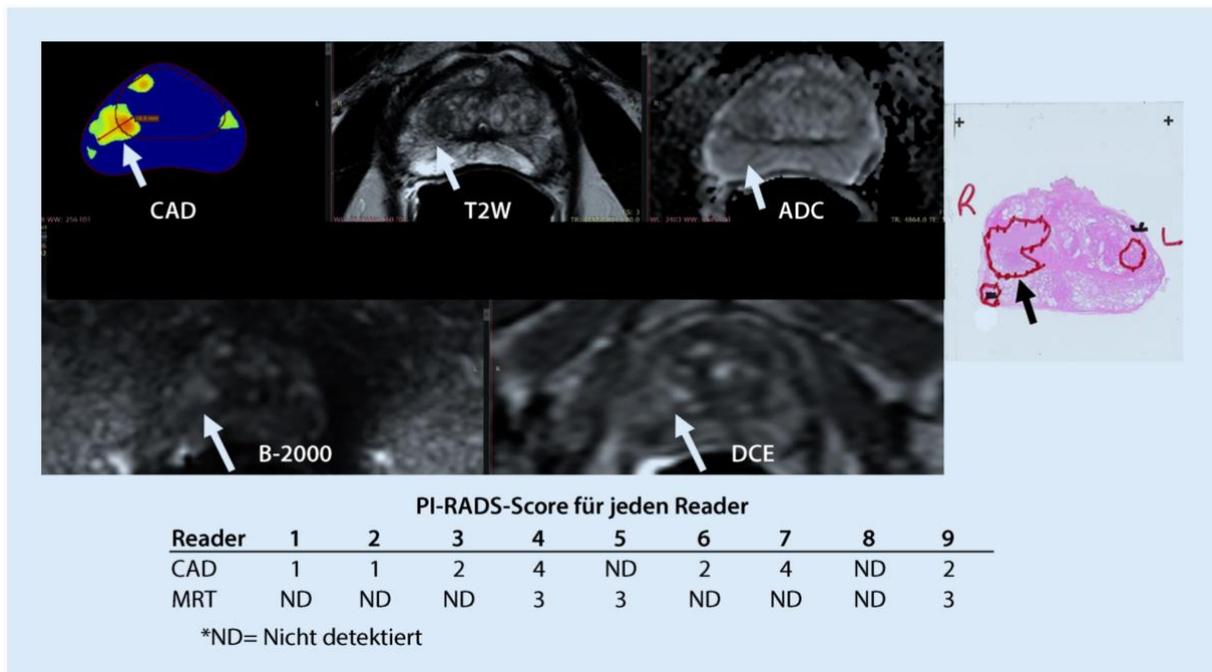


Abb. 5: mpMRT-Aufnahme der Prostata mit PCa-verdächtiger rechtsapikaler Läsion in PZ und TZ (Pfeile). „CAD“ = Computer-aided diagnosis. Die Tabelle zeigt, dass ohne CAD-Unterstützung 6 von 9 Radiologinnen und Radiologen die Läsion nicht erkannten. Mithilfe des CAD erkannten 7 von 9 Radiologinnen und Radiologen die Läsion. Histologisch konnte ein PCa mit GS 4 + 4 nachgewiesen werden [133, 257]. Mit freundlicher Genehmigung von: Springer Nature / Der Radiologe 60, 48-55 [Artificial intelligence and radiomics in MRI-based prostate diagnostics]. Hamm, C.A., Beetz, N.L., Savic, L.J., and Penzkofer, T. (2020).

## 2.12. Entwicklung eines Machine-Learning-Algorithmus: Radiomics Pipeline

Um aus radiologischen Rohdaten radiomische Features zu generieren und diese zur Prädiktion von Zielparametern verwenden zu können, erfolgt ein Mehrschrittprozess, auch bezeichnet als „Radiomics-Pipeline“ [207]. Jeder Schritt dieser Pipeline bringt spezifische Probleme mit sich, die einzeln und in ihrer Summe die Reproduzierbarkeit der Ergebnisse einschränken können. Daher sollten die einzelnen Schritte möglichst standardisiert durchgeführt werden [258-260]. Die Merkmale, die in Kombination prädiktiv für einen bestimmten Zielparameter sind, werden als „Radiomics-Signatur“ bezeichnet [204, 207].

Eine Radiomics Pipeline umfasst klassischerweise die Abfolge folgender Arbeitsschritte:

### 2.12.1. Datenakquisition

Die Erstellung einer bildgebenden Untersuchung stellt den ersten Schritt in der Pipeline dar. Hierbei spielen zum Beispiel Herstellerunterschiede der verwendeten Geräte eine Rolle, die die Ergebnisse von Radiomics-Auswertungen beeinflussen können; dies hat eine Studie zur Interscanner-Variabilität in der CT-Bildgebung für Radiomics ergeben [261]. Auch die Untersuchungsmodalität und Sequenzen sind bedeutsam; so konnte eine weitere Studie erhebliche Unterschiede im Hinblick auf die Robustheit bzw.

Reproduzierbarkeit von Radiomics Features in den verschiedenen MRT-Sequenzen demonstrieren [262].

### 2.12.2. Segmentierung bzw. Region of Interest-Auswahl

Segmentierung bezeichnet die Delineation von anatomischen Strukturen und von relevanten Bildregionen wie Läsionen zur Erstellung einer ROI. Auf Basis der segmentierten ROIs erfolgt die weitere Bildanalyse.

Im Falle der Prostata kann eine Einteilung des Organs z.B. in periphere Zone und Transitionalzone erfolgen, und eine fragliche Läsion gezielt abgegrenzt werden. So wird als Grundlage für die weitere Analyse klar vorgegeben, welche Bereiche in einer MR-Aufnahme welcher anatomischen Region oder einer Läsion zuzuordnen sind.

Herkömmlicherweise wird die Segmentierung manuell durchgeführt: die Konturen einer bestimmten Bildregion werden in zweidimensionale Schichtbilder eingezeichnet. Hierzu liegen allerdings bis dato noch keine Standards vor. Die manuelle Segmentierung bringt auch Nachteile mit sich. So unterliegt die Qualität der Delineationen den Limitationen des menschlichen visuellen Systems und ist zudem untersucherabhängig. Der Inter-Reader Bias kann problematisch im Hinblick auf die Reproduzierbarkeit der Ergebnisse sein. So wies eine Studie, in der Läsionssegmentierungen von sechs erfahrenen Radiologinnen und Radiologen an Prostata mpMRT-Aufnahmen von 20 Patienten durchgeführt wurden, eine durchschnittliche Abweichung von 0,23 cm zwischen den eingezeichneten Segmentierungen der einzelnen Befunderinnen und Befunder nach. Darüber hinaus wurden in dieser Studie 69 Satellitenläsionen, die insgesamt bei den Patienten vorlagen, von sämtlichen befundenden Ärzten nicht segmentiert, da sie nicht gesehen wurden [263].

Darüber hinaus ist die manuelle Segmentierung mit einem erheblichen Zeitaufwand verbunden. Sie muss je nach Fragestellung für jeden Patienten und für jede einzelne Schicht jeder einzelnen Sequenz einer umfangreichen bildgebenden Untersuchung wie einer mpMRT neu erarbeitet werden.

Zum Inter-Reader Bias kommt ein Intra-Reader Bias hinzu, da die manuelle Segmentierung „nach Augenmaß“ erfolgt und daher Schwankungen unterworfen sein kann.

Vor diesem Hintergrund wird das Potential einer automatisierten Prostata- und Läsionssegmentierung für den klinischen Alltag deutlich [202]. Es gibt alternativ zum manuellen Ansatz verschiedene Verfahren zur Generierung der ROIs, zum Beispiel die sogenannte Bounding Box, bei der eine rechteckige ROI über die Läsion gelegt wird, was allerdings auch umliegendes Gewebe in die Auswertung miteinschließt.

Mittels solcher Tools kann die aufwändige manuelle Segmentierung vereinfacht und standardisiert werden. Laut Kleesiek et al. können automatisierte Segmentierungen (z.B. durch KI-Methoden) die Konsistenz der Resultate verbessern, indem sie die potentiell erhebliche und für die Analysen nachteilige Intra- und Interreader-Variabilität verhindern [264].

Eine automatisierte Segmentierung hat unter diesen Gesichtspunkten zwar Vorteile, jedoch muss auch hier auf eine ausreichende Genauigkeit und mögliche Fehler eines automatisierten Vorgangs geachtet werden, da die Segmentierungen einen grundlegenden Schritt in der Machine Learning-Pipeline und somit eine potentielle (Folge-) Fehlerquelle darstellen können. Das Open-Source Tool nnU-Net, ein Verfahren zur automatischen Konfiguration selbstständig lernender Algorithmen, hat

sich in diesem Kontext als ein sehr erfolgreicher Ansatz zur Segmentierung herausgestellt [265]. Eine weitere Studie wendete mit Erfolg ein KNN (3D U-Net) zur Läsionssegmentierung der Prostata an [266].

Eine Studie konnte eine automatisierte DL-Methode zur Segmentierung der Prostata in der MR-Bildgebung entwickeln, die sich als akkurater erwies als die manuelle Segmentierung [267]. Yan K et al. konnten ein Modell entwickeln, das durch Extraktion und Kombination von Multi-Level-Features in einem tiefen Netzwerk eine verlässliche Segmentierung der Prostata in MR-Bildern erreichte [268].

Auch scheint es aussichtsreich, PZ und TZ getrennt voneinander bzw. zonenspezifisch auf unterschiedliche bildgebende Eigenschaften von PCa und gesundem Prostatagewebe zu betrachten. Eine Studie von Antonelli et al. demonstrierte, dass ein PZ/TZ-zonenspezifisch trainierter ML-Algorithmus Läsionen mit einem GS = 4 mit größerer Wahrscheinlichkeit korrekt vorhersagen konnte als erfahrene Radiologinnen und Radiologen. Hierzu wurden neben einem selektierten Feature-Set auch klinische Parameter wie PSAD und Drüsenvolumen herangezogen [269]. Dass die zonale bzw. histologische Gliederung der Prostata Einfluss auf die Performance von Machine Learning-Algorithmen hat, konnte auch in einer anderen Studie gezeigt werden. Zwar wurde kein Unterschied beim Vergleich eines TZ-spezifischen mit einem zonunenabhängigen Algorithmus gezeigt, beim PZ-spezifischen Klassifikator hingegen konnte ein signifikanter Unterschied gezeigt werden [270]. Laut einer weiteren Studie ist die Performance des Systems für PZ-Läsionen am besten [257].

### 2.12.3. Datenvorverarbeitung („Preprocessing“)

Preprocessing beschreibt das Vorbereiten von Daten für weitere, z.B. statistische Analysen. Das Ziel des Preprocessing ist das Aufbereiten von Rohdatensätzen zur weiteren Verarbeitung, ohne die enthaltenen Informationen stark zu verändern und somit Bias miteinzuführen und das Analyseergebnis zu verfälschen.

Das Daten-Preprocessing ist ein äußerst wichtiger Einflussfaktor auf die Ergebnisse einer Radiomics- bzw. Machine Learning-Analyse, da diese insbesondere Robustheit und Reproduzierbarkeit von Radiomics Features in hohem Maße beeinflussen können [271].

Darüber hinaus bestimmt die Datenqualität die Lernfähigkeit eines Radiomics-basierten ML-Modells entscheidend und Effizienz und Diagnosegenauigkeit eines Modells sind hiervon abhängig [272].

Aufgrund der technischen Gegebenheiten bei MRT kann es zu Störfaktoren bzw. Artefakten kommen; hierzu zählen z.B. Bildrauschen, Blurring- bzw. Unschärfeartefakte, Trunktionsartefakte an Grenzflächen von Geweben mit sehr unterschiedlichen Signalintensitäten, Partialvolumeneffekte (Aufreten mehrerer Gewebeintensitäten innerhalb eines Voxels bei zu geringer räumlicher Auflösung) usw. Somit sind die Bilddaten nicht immer direkt zur anschließenden Bildanalyse geeignet. Zur Datenvorverarbeitung gehört auch die Datenbereinigung (z.B. die Korrektur oder das Entfernen beschädigter oder unvollständiger Daten oder Duplikate, und die Bereinigung von Ausreißern und Inkonsistenzen).

Auch gehört der Histogrammausgleich (Histogramm: graphische Darstellung von Häufigkeitsverteilungen von (Bild-) Werten) zu den Standardmethoden der digitalen Bildverarbeitung. Hiermit wird der Kontrast und damit die visuelle Qualität von (Graustufen-) Bildern optimiert. Der Histogrammausgleich kann z.B. mittels Python-Bildverarbeitungsbibliotheken durchgeführt werden. Die Binarisierung als extreme

Form der Histogramm-Skalierung wird bei der Segmentierung genutzt: durch eine Reduktion der Grauwerte eines Bildes auf 2 können Objekte voneinander abgetrennt und ROIs konstruiert werden.

Bei der Datenreduktion werden Redundanzen entfernt.

Zum Schritt der Datentransformation gehört die Normalization bzw. Datennormalisierung, die ein grundlegendes Element in der Datenverarbeitung beschreibt. Dabei werden Daten in zur weiteren Verarbeitung geeigneten Formate transformiert und redundante Daten entfernt, sodass eine Normalverteilung im Sinne einer Gauß'schen Glockenkurve gegeben ist. Standardisierte (Intensitäts-) Normalisierungstechniken als Schritt im Data-Preprocessing sind Gegenstand der Forschung, da unter anderem hiervon die Reproduzierbarkeit von Features stark abhängig ist [273, 274]. Besonders Texturanalysemethoden sind empfindlich für Inkonsistenzen, beispielsweise in der Datenakquisition (z.B. durch die Verwendung unterschiedlicher Scanner oder Untersuchungsprotokolle). Die Reproduzierbarkeit und Stabilität von aus MRT abgeleiteten Radiomics Features unterliegt diesen limitierenden Faktoren, weswegen an die Abläufe der Datenverarbeitung hohe Ansprüche gestellt werden müssen. Insbesondere hier kommen Normalisierungsmethoden zum Tragen, die Unterschiede in Datenakquisition und Rekonstruktion ausgleichen können und damit einen positiven Einfluss auf die Reproduzierbarkeit von Radiomics Features haben können [225, 275].

Eine weitere klassische Bildverarbeitungsoperation stellt die Anwendung von Filtertechniken dar. Hierbei sind Glättungsfilter (ein auf Mittelwertbildung basierendes Weichzeichnen von Bildern), Filter zur Erhöhung der Bildschärfe (Hervorhebung hochfrequenter Anteile in einem Bild; dies führt jedoch zu einer Verstärkung von Bildrauschen) und Kantenfilter (Hervorheben starker Helligkeitsgradienten in einem Bild; diese sind sehr anfällig für Bildrauschen) zu erwähnen. Der sogenannte Laplacian of Gaussian (LoG) ist hierbei ein erwähnenswerter Kantenfilter, da hierbei eine Kombination aus Glättung und Kantenfilter angewandt wird.

Daneben werden in der Signalverarbeitung Wavelet-Filter (Wavelets: „kleine Wellen“) verwendet, um Informationen aus einem Signal zu extrahieren und kann ebenfalls zur Reduktion von Bildrauschen dienen [276].

Daten-Preprocessing kann durch bestimmte Tools und Bibliotheken vereinfacht durchgeführt werden. Die open-source Python Plattform PyRadiomics v3.0 aus dem Computational Imaging & Bioinformatics Lab der Harvard Medical School bietet die Anwendung verschiedener Preprocessing Filter an, wie z.B. Wavelet und LoG zur Datenvorverarbeitung vor der Extraktion von Radiomics Features [277].

Obwohl diese Schritte des Preprocessings noch nicht Gegenstand der Image Biomarker Standardisation Initiative (IBSI) sind, einer unabhängigen internationalen Organisation, die an einer Standardisierung von bildgebenden Biomarkern arbeitet, konnten mehrere Studien zeigen, dass die Anwendung von Preprocessing Filtern (insbesondere Wavelet und LoG) hoch prädiktive Features hervorbringen [278-280].

#### 2.12.4. Datenaugmentierung und Balancing

Data-Augmentation oder Datenaugmentierung beschreibt eine Methode zum Erweitern eines Datensatzes in Situationen, in denen eine größere Menge an Trainingsdaten erforderlich wäre, als vorhanden ist. Auch zum Ausgleich von Imbalancen wird die Augmentierung angewendet, wenn zwischen Trainingsdatensätzen ein das Testergebnis verzerrendes zahlenmäßiges Ungleichgewicht an enthaltenen Gruppen vorliegt (in

dieser Studie die Gruppen „kein PCa“ (bzw. „benigne Läsion“), „klinisch nicht-signifikantes PCa“, und „klinisch signifikantes PCa“). Für ein valides Ergebnis sollten die Gruppen möglichst gleich verteilt, bzw. keine Beispiele unterrepräsentiert sein.

Um das häufige Problem dieses sogenannten Klassenungleichgewichts im Klassifizierungsdatensatz zu lösen, eignen sich verschiedene Techniken. Die simpelste Methode wäre ein einfaches Duplizieren von Daten, womit dem Modell jedoch keinerlei neue Informationen zugeführt werden würden. Mit Techniken wie z.B. Größenanpassung durch Interpolation, Datennormalisierung, Concatenation (Verkettung) etc. lassen sich anhand der vorhandenen Bilder geringfügig veränderte zusätzliche Bilder erzeugen, die den Datensatz ergänzen und das Klassenungleichgewicht ausgleichen („Balancing“). Außerdem lässt sich beispielsweise durch verschiedene Skalierungen (Zoom), Rotation oder Spiegeln („Flipping“) der ROIs die Anzahl der Datensätze zum Training vergrößern. Die Rotation kann hierbei um jeden Wert bis zu 360° erfolgen, das Flipping entweder um die horizontale oder die vertikale Achse. Hierbei wird im Einzelfall entschieden, welche Anwendung am sinnvollsten ist.

Eine weitere Methode stellen Sampling-Methoden dar. Hier bieten sich zum Ausgleich von Klassen-Imbalancen zwei Wege: Das Undersampling der Mehrheitsklasse, bei dem Beispiele aus der Mehrheitsklasse entfernt werden, oder ein Oversampling der Minderheitsklasse, bei dem Beispiele der Minderheitsklasse dupliziert oder neue Beispiele anhand der vorhandenen erstellt werden. Eine weitläufig angewandte Oversamplingmethode ist SMOTE (Synthetic Minority Oversampling Technique) [281]. Die Analyse mittels SMOTE setzt sich aus den ursprünglich vorhandenen und zusätzlich generierten Minderheitsstichproben zusammen. SMOTE entnimmt zufällige Stichproben aus der Minderheitsklasse, lokalisiert deren Nachbarwerte und wählt zwischen diesen beiden Werten ebenfalls zufällige Punkte aus. Die neu konstruierten Stichproben haben andere Merkmale als die ursprünglichen und gleichen so die Klassen-Imbalancen aus.

Die Schritte der Datenaugmentierung können erhebliche Auswirkungen auf die Ergebnisse einer Radiomics-Analyse haben [282].

Hierzu eignen sich zum Beispiel Funktionen von Radiomics-Softwarepaketen [283, 284].

#### 2.12.5. Extraktion von Radiomics Features

Die Feature-Extraktion erfolgt durch verschiedene mathematische Verfahren, die die Radiomics Features definieren und algorithmisch umsetzen. Dazu eignen sich z.B. quelloffene Programme, mit deren Hilfe verschiedene Features extrahiert werden können, wie dem oben erwähnten PyRadiomics. PyRadiomics kann verschiedene Feature-Klassen anhand gefilterter oder ungefilterter Daten extrahieren [284-286].

#### 2.12.6. Selektion von Radiomics Features

Nach der Extraktion müssen die Bildmerkmale mit der besten Aussagekraft selektiert werden. Im Normalfall werden hunderte radiomische Merkmale von den Software-Tools errechnet. Das sogenannte Overfitting stellt jedoch bei Radiomics-Analysen eine

potentielle Gefahr dar. Häufig steht einer nicht allzu hohen Anzahl von Trainingsbeispielen eine sehr hohe Anzahl von Features gegenüber. Werden zu viele Variablen vom Modell in die Analysen einbezogen, kommt es zu einer Überanpassung und das trainierte Modell liefert bei der Anwendung auf neue Daten unzureichende Ergebnisse.

Daher ist eine zahlenmäßige Reduktion der Features nötig, um zu vermeiden, dass Features berücksichtigt werden, die in Bezug auf die Fragestellung keine oder eine zu geringe Relevanz haben. Umgekehrt könnten bei einem zu einfachen Modell gegebenenfalls wesentliche Variablen nicht berücksichtigt werden. Daher bestimmt die richtige Auswahl der geeigneten Features in entscheidendem Maße das Ergebnis („Feature-Reduktion“) [287, 288].

Für die Feature-Selektion kommen beispielsweise maschinelle Lernverfahren zum Einsatz, die Features nach möglichst hoher Stabilität oder möglichst geringer Interkorrelation untereinander auswählen [219, 287]. Die interkorrelierten Features können identifiziert Features mit redundantem Informationsgehalt sollten ausgeschlossen werden [289, 290]. Hierzu können ML-Algorithmen verwendet werden, wie z.B. der Minimum Redundancy Maximum Relevance-Algorithmus (mRMR), der Features selektiert, die eine hohe aufgabenbezogene Relevanz haben, während sie untereinander möglichst wenig korreliert sind („Redundancy“) [291].

#### 2.12.7. Training des Modells

Um ein auf spezifische Fragestellungen anwendbares und übertragbares Modell zu entwickeln, muss dieses mit Beispielen trainiert werden. Je akkurater das Modell Klassifizierungen vornimmt, desto besser hat es anhand von Trainingsdaten gelernt. Der Prozess des iterativen Anpassens eines Algorithmus bzw. eines Modells auf diese Daten stellt den „fitting process“ oder auch das Training dar.

Die Eigenschaften des Zielparameters bestimmen die Auswahl des Radiomics-Modells. Es können verschiedene Modelle eingesetzt werden [207]. Diese werden im Abschnitt 3.1.3 „Entwicklung eines Radiomics-basierten Machine Learning Algorithmus: Klassifikationsmodelle“ näher erläutert.

Anhand von Trainingsdaten wird das Modell dann so lange optimiert, bis es möglichst gute Resultate erreicht.

#### 2.12.8. Interne Validierung des Modells

Validierungsdaten werden benötigt, um zu verhindern, dass Modelle entstehen, die nur auf dem ausgewählten Datensatz gute Ergebnisse erzielen, jedoch nicht bei neuen Daten. Im Kontext von Radiomics werden zu diesem Zweck zwei völlig voneinander unabhängige Datensätze empfohlen: einer zur internen Validierung (Training und Kreuzvalidierung (Cross-Validation, CV)) und einer zur externen Validierung (i.e. Prüfen von Performance und Generalisierbarkeit). Die Trainingsbeispiele sollten grundsätzlich von den Testbeispielen getrennt bleiben. Sowohl Trainings- als auch Testdatensatz sollten balanciert sein [292].

Bei einem von vornherein zahlenmäßig beschränkten Datensatz kann diese Aufteilung in Trainings- und Testdatensatz zu einer problematischen Verringerung der jeweiligen

Datenvolumina und damit zu instabileren Ergebnissen führen. Um dennoch eine effiziente Validierung zu gewährleisten, bietet sich die Methode der Kreuzvalidierung an. Steht zur Validierung kein separater Datensatz zur Verfügung, können wiederholt Stichproben aus den vorhandenen Datensätzen gezogen werden, um somit „neue“ Datensätze zu generieren. Bei der 5-fachen Kreuzvalidierung werden fünf Teildatensätze gebildet. Vier davon dienen jeweils zur Modellentwicklung, der fünfte als Validierungs-Datensatz. Bei diesem Verfahren wird also ein an jeweils vier Teildatensätzen trainiertes Modell auf den verbleibenden Teildatensatz angewandt und das Ergebnis jeweils mit den tatsächlichen Zielwerten verglichen. Die Kreuzvalidierung als interne Validierungsmethode wird demnach zur Modell-Optimierung angewendet. Sie gilt als eine Standard-Validierungsmethode [293].

#### 2.12.9. Test und externe Validierung des Modells

Besonders die Anwendung der Kreuzvalidierung während der Algorithmusentwicklung kann zu überoptimistischen Einschätzungen der Modellgüte führen. Aus diesem Grund ist ein unabhängiger Testdatensatz wünschenswert, um die Generalisierbarkeit des Modells (und die Übertragbarkeit in ein klinisches Setting) auf neue Daten einschätzen zu können [201].

Die (Kreuz-) Validierung sollte daher möglichst durch einen zusätzlichen Test auf einem zum Training und ggf. zur Kreuzvalidierung nicht genutzten Datensatz ergänzt werden, bei dem die Generalisierbarkeit des Modells untersucht wird. Idealerweise wird das trainierte Modell zudem auf an anderen Zentren erhobenen Daten angewandt, um auch die Übertragbarkeit auf andere Standorte zu überprüfen.

#### 2.12.10. Performance – Maße

Beim „Data Mining“ (automatisierte Verarbeitung großer Datenmengen durch computergestützte Methoden zur Ableitung von aussagekräftigen Informationen wie Mustern, Zusammenhängen, Trends usw.) werden oft mehrere Modelle verglichen und dasjenige mit der besten Performance ausgewählt. Dazu werden vergleichbare Kennzahlen benötigt. Die Auswahl der geeigneten Metrik ist bei der Einordnung der Performance von Machine Learning Modellen entscheidend.

Die Ergebnisse einer binären KI-Klassifikation werden üblicherweise in einer sogenannten Confusion Matrix dargestellt, einer 2D-Anordnung der Klassifikationsergebnisse. Die Confusion Matrix dient zur Evaluation einer KI-Anwendung. Hier wird nur in „richtige“ und „falsche“ Klassifikationen unterschieden, genauer gesagt in die Elemente „korrekt positiv“, „korrekt negativ“, „falsch positiv“ und „falsch negativ“. Aus der Confusion Matrix lassen sich weitere Größen zur Leistungsdiagnostik für ein Klassifikationsmodell ableiten. Eine wichtige Performance-Metrik ist die Klassifizierungsgenauigkeit („Accuracy“). Sie gibt den Anteil korrekter Prädiktionen an der Gesamtheit der Prädiktionen an und berechnet sich mit der Formel:

$$\text{Accuracy} = \frac{\text{richtig positiv} + \text{richtig negativ}}{\text{Anzahl der Datensätze}}$$

Die Accuracy kann einen maximalen Wert von 1 betragen. Je näher der Wert an 1, desto genauer ist der Klassifikator.

Jedoch ist die Accuracy eine sehr einfache Metrik, die insbesondere bei binären Klassifikationen an unausgeglichenen Datensätzen problematisch sein kann: wenn ein Klassifikationsmodell alle Beispiele aus dem überrepräsentierten Datensatz einer Klasse zuordnen würde, würde er eine hohe Genauigkeitsrate erreichen, die jedoch nicht sinnvoll wäre. Daher sollten weitere Metriken verwendet werden. Die Präzision („Precision“) ist eine klassenspezifische Leistungsmetrik, die den Anteil der richtig positiven an allen positiven Klassifikationen bezeichnet.

Der sogenannte Rückruf („Recall“) bezeichnet einen Bruchteil der Stichproben einer Klasse, die vom Modell richtig klassifiziert wurden (richtig positive / (richtig positive + falsch negative)).

Oft sind je nach Anwendung sowohl Precision als auch Recall sinnvolle Größen, die mit dem sogenannten F1-Score als harmonisches Mittel kombiniert angegeben werden können.

Eine häufig angewendete Metrik zur Darstellung der Güte eines Modells ist die ROC („Receiver-Operating-Characteristic“) -Kurve. Insbesondere im Bereich des Machine Learning wird die ROC eingesetzt, um die Performance eines Klassifikators mit einer dichotomen Zielvariable zu evaluieren [294].

Hierbei werden auf der y-Achse die relativen Häufigkeiten der richtig positiv klassifizierten Fälle gegen die relativen Häufigkeiten der falsch positiv klassifizierten Fälle auf der x-Achse aufgetragen. Die entstehende Kurve schneidet die Werte (0/0) und (1/1). Die zufällige Zuordnung zu einer Klasse (Raten) entspricht einer Winkelhalbierenden: Für jeden möglichen Grenzwert zur Vorhersage einer Zielvariable gibt es gleich viele richtig positive und falsch positive Ergebnisse. Je zuverlässiger ein Modell Vorhersagen trifft, desto mehr weicht also die ROC von der Winkelhalbierenden ab. Die AUC („Area under the curve“) bezeichnet die Fläche unter der ROC-Kurve. Im Fall der Winkelhalbierenden hat diese aufgrund der Verteilung der Zufallswerte eine Fläche von 0,5 Einheiten. Dies entspricht also dem schlechtesten Ergebnis. Ein optimaler Test, der immer die korrekte Zielvariable vorhersagt, hat einen AUC-Score oder eine AUC von 1. Dementsprechend sind logistische Regressionsmodelle umso zuverlässiger und besser, je näher ihre AUC bei 1 liegt. Wird also ein Patient mit nicht signifikantem und ein Patient mit klinisch signifikantem PCa betrachtet, entspricht die AUC genau der Wahrscheinlichkeit, mit der die Patienten der jeweils richtigen Gruppe korrekt zugeordnet werden.

Ein Modell arbeitet an einem wählbaren „Arbeitspunkt“ der ROC-Kurve, durch den sich eine Sensitivität, Spezifität und Genauigkeit ergibt. Durch Anpassung dieses „Thresholds“, zum Beispiel im Sinne einer hohen Sensitivität, ergibt sich eine konkrete Wahrscheinlichkeit für eine richtige bzw. falsche Zuordnung. Die Wahl des Arbeitspunktes ist aufgabenabhängig.

Ein Vorteil am AUC-Score liegt darin, dass sich mit ihm die Performance verschiedene Modelle direkt vergleichen lässt.

Abb. 6 veranschaulicht die beschriebenen, klassischen Schritte einer Radiomics-Pipeline.

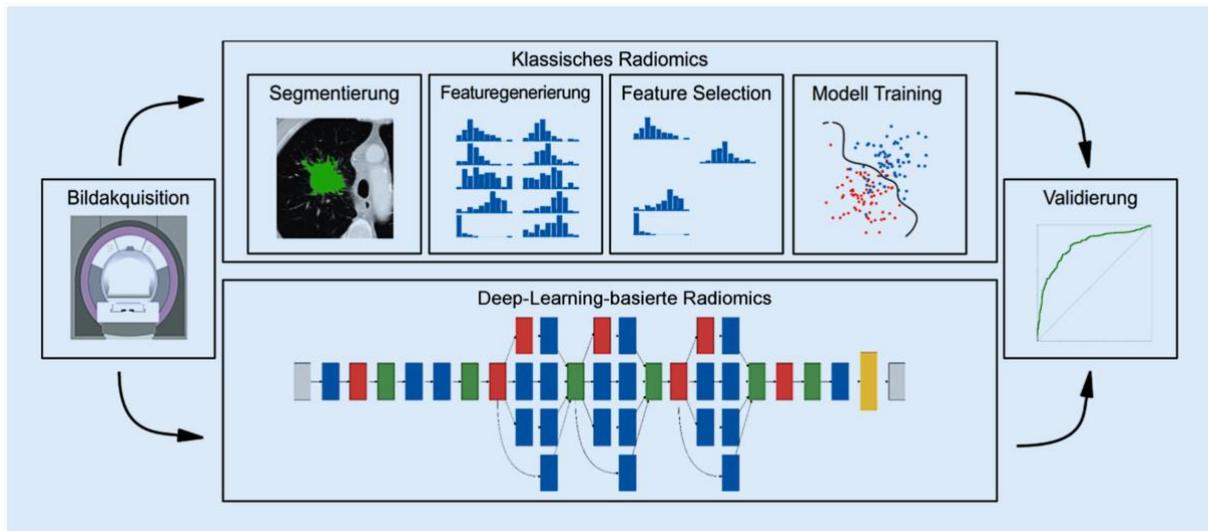


Abb. 6: Ablauf der Arbeitsschritte in einer üblichen Radiomics Pipeline am Beispiel einer CT-Aufnahme der Lunge. Es werden zwei Verfahren gezeigt: der obere Pfeil nach der Bildakquisition führt zum klassischen Verfahren einer Pipeline, der untere Pfeil zum Deep Learning-Verfahren (Erläuterung in Abschnitt 2.14) [201].

Mit freundlicher Genehmigung von: Springer (Der Pathologe; (Radiomics-AI-based image analysis, Demircioğlu, A)), 2019.

### 2.13. Problematik der Reproduzierbarkeit und Generalisierbarkeit von MRT-basierten Features

Radiomics und Machine Learning bergen vielversprechendes Potential im Hinblick auf ihre Anwendungsmöglichkeiten. Jedoch liegt die Etablierung von Radiomics-Analysen in der klinischen Praxis noch in weiter Ferne. Es müssen einige Problemfelder berücksichtigt werden: die Komplexität des klassischen Radiomics-Workflows bringt als zentraler Punkt eine schlechte Wiederholbarkeit (Erreichen gleicher Ergebnisse bei Verwendung derselben Untersuchungsprotokolle und Materialien), Reproduzierbarkeit (Erreichen gleicher Ergebnisse durch Anwendung variierender Untersuchungsprotokolle und Materialien) und damit Generalisierbarkeit (Übertragbarkeit und Anwendung von Modellen an z.B. verschiedenen Kliniken) mit sich.

Gerade bei mpMRT-Daten als Input, die qualitative Bildgebungen (Sequenzen wie T<sub>1</sub>- und T<sub>2</sub>-Wichtungen, FLAIR (fluid attenuated inversion recovery) usw.) einbezieht, sind Signalintensitäten von Faktoren wie Bildakquisition an unterschiedlichen Scannern oder unterschiedlichen Rekonstruktionsalgorithmen abhängig und damit Schwankungen unterworfen, denen gegenüber sich Radiomics Features nicht robust verhalten [295-297]. Die Robustheit der verwendeten Features muss demnach evaluiert werden, ebenso wie die Entwicklung komplexer Vorhersagemodelle standardisiert werden muss [298]. Mehreren Studien zufolge sind entsprechend nur einige Radiomics Features sowohl in Test-Retest-Studien wiederholbar als auch mit anderen Studienparametern reproduzierbar [295, 299, 300]. Im Umkehrschluss allerdings kann eine gewisse Diversität im Hinblick auf die Bildakquisition bei der Entwicklung von robusten Radiomics-Signaturen von Vorteil sein [275].

Wie beschrieben, konnten einige Studien bereits interessante Radiomics-Anwendungen in verschiedenen Bereichen zeigen. Allerdings gibt es bis dato keine festgelegten Standards oder einheitlichen Qualitätskriterien für Radiomics-Workflows und die Entwicklung von Machine Learning-Modellen. Eine Standardisierung von Input-Daten (z.B. standardisierte

Untersuchungsprotokolle, Kriterien für die Bildqualität, die Verwendung verschiedener Scanner), der Methodik der Segmentierung (z.B. die Verwendung von manuellen Segmentierungen mehrerer Befunderinnen und Befunder), von Evaluationskriterien usw. würde sich positiv auf die Reproduzierbarkeit der Analyseergebnisse auswirken.

Es gibt bereits mehrere Ansätze zur Einordnung der Studienqualität zu Radiomics. Eine Möglichkeit zur Einschätzung der methodologischen Qualität von Studien ist der Radiomics Quality Score (RQS), der zwar noch keine weitläufige Anwendung findet, aber bereits zum Assessment von Studien zu Radiomics in der onkologischen Bildgebung herangezogen wurde und wird [301]. Auch die Checklist for Artificial Intelligence in Medical Imaging (CLAIM) soll eine Vereinheitlichung der Methodologie vorantreiben [302]. Die IBSI gibt Empfehlungen für das Preprocessing und die Feature-Extraktion [278]. Um die Ergebnisse und deren Darstellung einheitlich und vergleichbar zu machen, gibt es die Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Checkliste.

Unter Anwendung von RQS und TRIPOD wurde in einer Studie die methodologische Qualität einiger Studien analysiert. Diese kam zu eher unbefriedigenden Ergebnissen [303]. Laut einem systematischen Review sind auch die Studien zu MRT-basierten Radiomics zur PCa-Diagnostik aufgrund qualitativer Unzulänglichkeiten in Bezug auf die Methodologie noch nicht geeignet, um in die klinische Anwendung übertragen werden zu können. Auch bei Fokus auf Studien von High-Impact Journals zu onkologischer Anwendung von Radiomics wurde nur ein durchschnittlicher RQS-Score von 26,1 % erreicht [304].

Eine verbesserte Generalisierbarkeit wäre durch eine Berücksichtigung der dargestellten Punkte realisierbar, womit wiederum die Implementation von Radiomics-Analysen in die klinische Praxis realistischer würde.

#### 2.14. Künstliches neuronales Netzwerk, Deep Learning und Faltungsnetz

Das sogenannte künstliche neuronale Netzwerk (KNN), ein Modell aus der Neuroinformatik, ist die gegenwärtig vielversprechendste Technik zur Anwendung an radiologischen Bildern. Die Theorie des KNN existierte bereits in den 1940er Jahren, die Entwicklung stagnierte jedoch bis in die 1980er Jahre aufgrund noch unzureichender Computer-Rechenleistungen [227].

Es ahmt in seiner biologischen Struktur und Funktionsweise die Neuronen eines Gehirns nach und folgt einem dreischichtigen Prinzip. Ein Knotenpunkt, auch Input-Neuron genannt, fungiert als Eingang („Input“), der Informationen aufnimmt. Diese werden in einer verborgenen Schicht („hidden layer“) verarbeitet, und als Ausgang („Output“) wird ein Ergebnis ausgegeben.

Jedem Neuron wird bei der Informationsverarbeitung ein individuelles Startgewicht zugewiesen. Somit gibt jedes Neuron der anliegenden Information ein individuelles Gewicht und gibt diese gewichtete Information an das nächste Neuron weiter, sodass letztlich ein „Pfad“ an Neuronen entsteht, den die Daten durch das Netzwerk genommen haben. Diese Berechnungen führen am Output-Neuron schließlich zu einem Ergebnis, wobei es zu Fehlern kommen kann. Entsprechend der Gewichtung der Neurone kann berechnet werden, wie maßgeblich ein Neuron an dem Fehler war. Im folgenden Lerndurchlauf kann dann eine Anpassung der Gewichte erfolgen, sodass der Fehler minimiert wird. Die Gewichtung der Eingabedaten wird anhand der Trainingsbeispiele erlernt und verbessert, um die Abweichung zur richtigen Ausgabe zu minimieren. Das Netzwerk lernt also anhand bereits existierender Outputs, die als „Ground Truth“ dienen [305].

Verschaltet man mehrere Neuronenschichten, erhält man ein künstliches neuronales Netz. KNN können in ihrer hidden layer unterschiedlich komplex sein: Zwischen den Inputneuronen und der Outputschicht des Algorithmus kann eine variable Anzahl von versteckten Netzwerkschichten liegen, welche aus Knotenpunkten in verschiedenen Strukturen bestehen.

Jeder Knotenpunkt ist mit Knotenpunkten in anderen Schichten verbunden, vergleichbar mit den Neuronen des Gehirns. Jedes „Neuron“ kann die Inputs aus unteren Schichten unterschiedlich miteinander kombinieren, ein komplexeres Output generieren, dieses an die nächste Schicht weitergeben und somit Schicht für Schicht Daten weiter abstrahieren [255, 306].

Beim Deep Learning (DL) verschaltet man mehrere dieser Schichten hintereinander. Je mehr Schichten enthalten sind, desto tiefer ist das Netz.

Während einfache netzwerkbasierende Algorithmen nur wenige Schichten aufweisen, beinhalten DL-Algorithmen eine größere Anzahl an Schichten [256, 307].

Im Gegensatz zu anderen ML-Verfahren, bei welchen die Merkmalsextraktion vom Menschen vorgegeben wird, werden beim KNN während des Trainings Merkmalsdetektoren ausgebildet, die autonom relevante Merkmale erkennen. Diese können hochkomplex sein und sich sogar dem menschlichen Verständnis entziehen [255]. KNN benötigen also im Gegensatz zu klassischen ML-Modellen weder vorher beigebrachtes Wissen noch durch den Menschen vorgegebene „handgemachte“ Features [308, 309]. Auch bei einem KNN tritt bei einer zunehmenden Anzahl an Parametern die Gefahr des Overfittings auf, sodass auch DL-Modelle analog zu anderen ML-Modellen kritisch evaluiert und validiert werden müssen.

DL-Methoden erzielen im Bezug auf Organ-Segmentierungen bereits die größten Erfolge und sind inzwischen sogar die dazu fast ausschließlich verwendete ML-Technik [310]. Auch zur Klassifizierung werden DL-Methoden bzw. KNNs bereits erfolgreich angewandt [256, 311].

Abb. 7 zeigt eine vereinfachte Form eines KNNs.

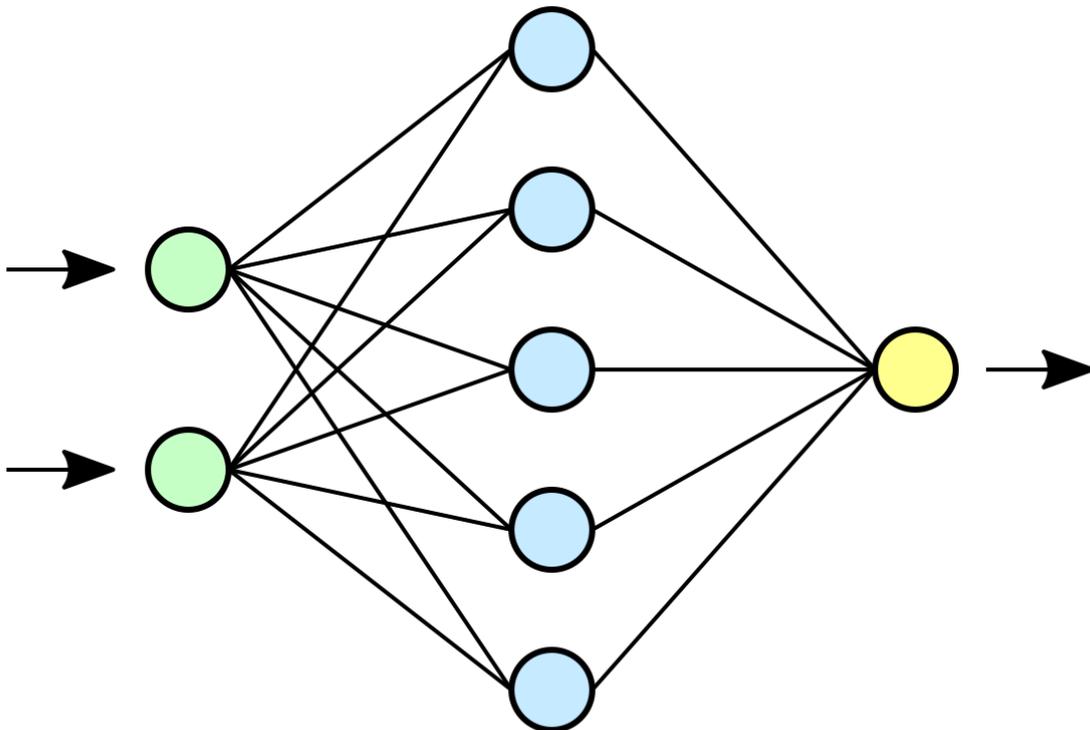


Abb. 7: Vereinfachte Darstellung eines neuronalen Netzwerks. Die Input- „Neuronen“-Schicht ist in grün, die Output-„Neuronen“-Schicht in gelb dargestellt. Blau stellt eine versteckte Schicht dar. Licensed under <https://creativecommons.org/licenses/by/1.0/>. Retrieved from: [https://upload.wikimedia.org/wikipedia/commons/3/3d/Neural\\_network.svg](https://upload.wikimedia.org/wikipedia/commons/3/3d/Neural_network.svg)

Es gibt verschiedene Arten solcher Netzwerke. Das Convolutional Neural Network (CNN) oder auch Faltungsnetz stellt eine Architektur für neuronale Netze dar, die insbesondere bei Bilderkennungsaufgaben und Klassifizierungen Anwendung findet. Die CNN-Struktur wurde 1994 erstmals beschrieben und ist seither einer der erfolgreichsten Algorithmen speziell im Bereich der Bilderkennung [312, 313].

Die Fähigkeit solcher DL-Modelle, unterschiedliche Datentypen als Inputs verwenden zu können, ist ein besonders interessanter Aspekt im Hinblick auf die heterogenen Daten, die das Gesundheitswesen generiert [307]. Dadurch stellt das DL insbesondere für die Anwendung in der Medizin und bei Vorhandensein von großen Datenmengen zum Training die Methode der Wahl dar.

## 2.15. Zielsetzung dieser Arbeit

Durch die Anwendung von Radiomics, also durch die integrative Analyse bildgebender und klinischer Parameter, soll ein Analysealgorithmus entwickelt werden, der es ermöglicht, Patienten mit signifikantem Prostatakarzinom noch vor Biopsie oder Prostatektomie zu identifizieren, d.h. bereits zum Zeitpunkt der Diagnosestellung soll somit eine Risikostratifizierung erfolgen. Hierzu sollen Methoden des Machine Learning bzw. Deep Learning auf mpMRT-Aufnahmen angewendet werden.

1. Es soll geprüft werden, ob mit einem Radiomics-basierten Analysealgorithmus eine Differenzierung zwischen signifikanten und nicht-signifikanten PCa möglich ist.
2. Hierzu soll im Gegensatz zu dem Läsions-basierten Ansatz (PI-RADS) ein schichtbasierter Ansatz verfolgt werden.
3. Es soll ein klassischer Radiomics-Ansatz mit einem Deep Learning-Ansatz verglichen werden.

### 3. MATERIAL UND METHODEN

#### 3.1. Ethikvotum

Für diese monozentrische, retrospektive Studie lag seitens der Ethik-Kommission II der Universität Heidelberg am 21.06.2018 ein positives Ethikvotum vor (Aktenzeichen 2018-846R-MA).

Diese Studie wurde nach Vorliegen der Bewilligung seitens der Ethikkommission begonnen.

#### 3.2. Patientenkollektiv

Zur Auswahl geeigneter Patienten wurden für diese Studie am Institut für Klinische Radiologie und Nuklearmedizin, Universitätsmedizin Mannheim (IKRN) der Medizinischen Fakultät der Universität Heidelberg Daten erhoben.

Es wurden folgende Parameter erhoben:

- Gleason-Score
- PSA-Wert zum Zeitpunkt der Diagnose
- PI-RADS-Score
- Prostatavolumen (in ml)
- Weitere klinische Parameter (DRU)
- Bildgebung: T<sub>2w</sub>, ADC, DWI, DCE

Spezifische Einschlusskriterien für die Erstellung des Patientenkollektivs waren:

- Einteilung in drei Gruppen nach: klinisch signifikantes PCa ( $GS \geq 4 + 3$ ), klinisch nicht signifikantes PCa ( $GS < 4 + 3$ ), kein PCa
- Gleason-Score
- Vorliegen von mindestens T<sub>1w</sub>-, T<sub>2w</sub>- und Diffusionsaufnahmen zum Zeitpunkt der Diagnose
- PSA-Wert

Spezifische Ausschlusskriterien:

- Keine vollständigen MRT-Datensätze vorliegend
- Relevante bzw. übermäßige Artefakte in MRT
- Keine Läsion in konsensueller Befundung
- Unvollständige oder fehlende Segmentierung

Nach diesen Kriterien erfolgte die retrospektive Identifikation von 405 Patienten. Im Zuge der Vorverarbeitung der Datensätze stellten sich teilweise unvollständige MRT-Datensätze oder Artefakte in der Bildgebung heraus; außerdem kam es durch technische Fehler bei der Segmentierung zu fehlerhaften oder unvollständigen Segmentierungen. Die fehlerhaften Dateien wurden aus der Studie ausgeschlossen. Das Flowchart in Abb. 8 gibt den Workflow des Patienteneinschlusses wieder. Von 405 eingeschlossenen Patienten lagen bei 326 mpMRT vor. Davon mussten 7 Patienten aufgrund eingeschränkter Bildqualität ausgeschlossen werden. Weitere 15 Patienten mussten aufgrund unvollständig oder fehlerhaft

segmentierter Datensätze ausgeschlossen werden, außerdem weitere 7, bei denen nach konsensueller Bildbefundung keine Läsion vorlag. Die finale Studienkohorte setzte sich aus 297 Patienten zusammen.

Das Vorliegen eines klinisch signifikanten PCa wurde bei einer Einschätzung als ISUP-Klasse 2 oder höher festgelegt, was einem GS von GS = 3 + 4 oder höher entspricht. Das klinisch nicht signifikante PCa wurde demnach durch einen GS = 3 + 3 oder weniger charakterisiert.

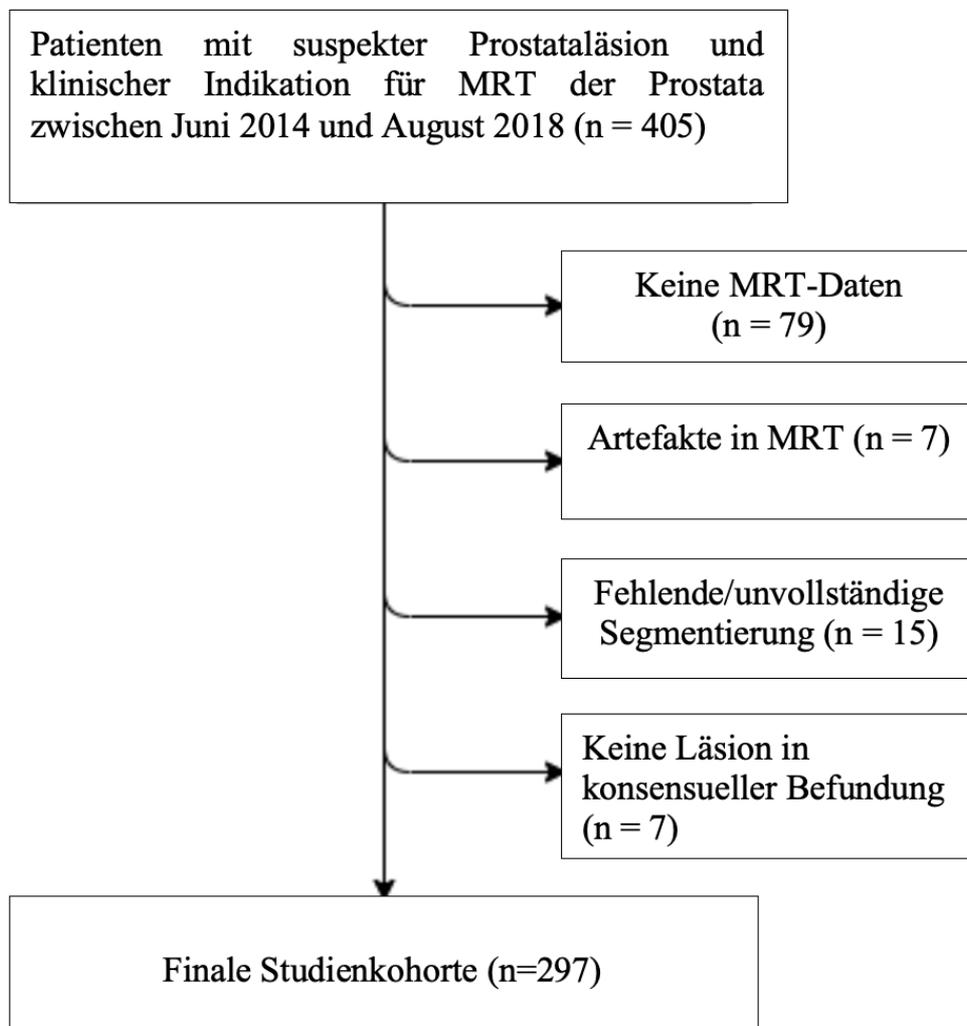


Abb. 8: Flowchart der Einschluss- und Ausschlusskriterien.

Die mpMRT-Aufnahmen wurden gemäß den Vorgaben für Untersuchungsprotokolle der PI-RADS v2 angefertigt. Tab. 3 gibt die Details der angewendeten mpMRT-Protokolle wieder. Die Untersuchungen wurden auf zwei 3T MRT-Geräten durchgeführt (Siemens Magnetom Skyra und Trio, Erlangen, Deutschland).

Tab. 3: Angewendete multiparametrische MRT-Protokolle der Prostata

| MRI sequences  | Orientation | TE (ms) | TR (ms) | Slice thickness (mm) | FoV read (mm) |
|--|-------------|---------|---------|----------------------|---------------|
| <b>MRI scanner no. 1 (Siemens, Magnetom Trio, 3T)</b>  |             |         |         |                      |               |
| DWI b50/500/800  | tra         | 88      | 5000    | 3                    | 204           |
| t2_haste_sag   | sag         | 105     | 1200    | 4                    | 281*300       |
| t2_tse_cor   | cor         | 101     | 4000    | 3                    | 200           |
| t2_tse_tra   | tra         | 101     | 4580    | 3                    | 200           |
| t2_tse_sag   | sag         | 101     | 4190    | 3                    | 200           |
| t1_vibe_cor_iso  | cor         | 2.15    | 4.71    | 0.91                 | 290           |
| t1_tse_tra   | tra         | 9.8     | 858     | 3                    | 200           |
| <b>MRI scanner no. 2 (Siemens, Magnetom Skyra, 3T)</b> |             |         |         |                      |               |
| DCE (Grasp)  | tra         | 1.69    | 3.85    | 1.5                  | 240           |
| DWI zoomit_epi   | tra         | 69      | 3800    | 3                    | 380           |
| DWI zoomit_epi_2000                                    | tra         | 91      | 4600    | 3                    | 380           |
| t1_twist_tra   | tra         | 1.97    | 5.05    | 3.5                  | 260           |
| t2_tse_tra   | tra         | 104     | 3300    | 3                    | 200           |
| t2_tse_sag   | sag         | 104     | 3600    | 3                    | 200           |
| t2_tse_cor   | cor         | 104     | 3200    | 3                    | 200           |
| DWI ep2d_dif   | tra         | 69      | 3600    | 4                    | 200           |

### 3.3.Referenzstandard: Histopathologische Korrelation

Als Referenzstandard für die klinische Signifikanz der PCa dienten die histopathologischen Befunde aus Fusionsbiopsie und systematischer Biopsie. Die histopathologische Korrelation bzw. das „Labeling“ erfolgte durch erfahrene Fachärztinnen und Fachärzte der Radiologie sowie Urologie.

Hierzu wurden die mpMRT als einzelne Schichten in Zusammenschau mit ihren PI-RADS-Scores und den GS bzw. histopathologischen Befunden der Biopsate beurteilt. Im Konsens wurden Labels für jede Zone (PZ, TZ) in jeder Schicht vergeben: „kein PCa“, „klinisch signifikantes PCa (csPCa)“, „klinisch nicht signifikantes PCa (non-csPCa)“. Auf allen ausgewählten Schichten war eine Läsion zu erkennen. Das Label „kein PCa“ bezeichnet daher in diesem Fall nicht eine unauffällige Prostata, sondern das Vorliegen einer als benigne eingestuften Läsion.

### 3.4.Segmentierung

Ziel der Segmentierung war es, die einzelnen Geweberegionen der Prostata (periphere Zone „PZ“, Transitionalzone „TZ“, und die Prostata als gesamtes Organ („Prostata gesamt“)) von den sie umgebenden Strukturen zu separieren und somit ROIs zu erzeugen, die im optimalen Fall ausschließlich Gewebe von einer der jeweiligen Segmentierungszonen beinhalten.

Zunächst wurden die für die Segmentierung relevanten Schichten der MRT-Aufnahmen jedes Patienten ausgewählt. Dazu wurden die kompletten Bilddatensätze mit dem Horos Open Source DICOM (Digital Imaging and Communications in Medicine) Viewer gesichtet (Horos Project, Genf, Schweiz, Version 3.3.6). Aus den T<sub>2</sub>-gewichteten MRT-Aufnahmen wurden diejenigen Schichten selektiert, auf denen Prostatagewebe erfasst war.

Die so ermittelten Schichten wurden anschließend zur Segmentierung im DICOM-Format in das Annotations-Tool LIFEx importiert (Nioche et al., 2018).

Es wurden nun entsprechend der Anatomie der Prostata die periphere Zone, die Transitionalzone, sowie die gesamte Prostata in jeder einzelnen der vorselektierten Schichten manuell eingezeichnet und als binäre Masken abgespeichert. Aus jeder segmentierten Bildschicht entstanden somit drei Binärbilder.

Dies erfolgte sowohl in den ausgewählten T<sub>2</sub>w-Aufnahmen als auch separat in den zugehörigen ADC-maps.

Abb. 9 und 10 zeigen exemplarisch den Vorgang der Segmentierung in PZ, TZ und das gesamte Organ sowie dasselbe in den zugehörigen ADC-maps an zwei repräsentativen Patientenfällen.

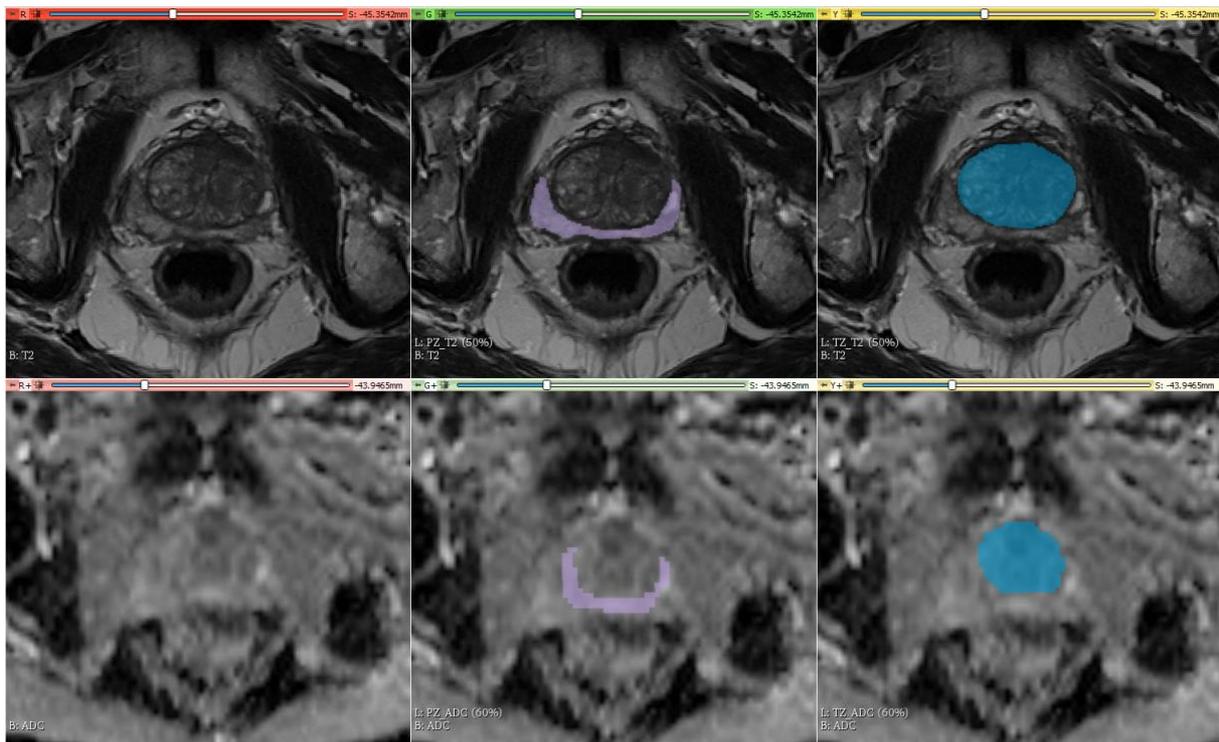


Abb. 9: Manuelle Segmentierung eines repräsentativen Falls (Gleason 4+3, PI-RADS 4, Läsion in TZ). Gezeigt sind die Delineationen der PZ (oben mittig) und TZ (oben rechts) in T<sub>2</sub>w, sowie der PZ (unten mittig) und TZ (unten rechts) in der ADC-map. .

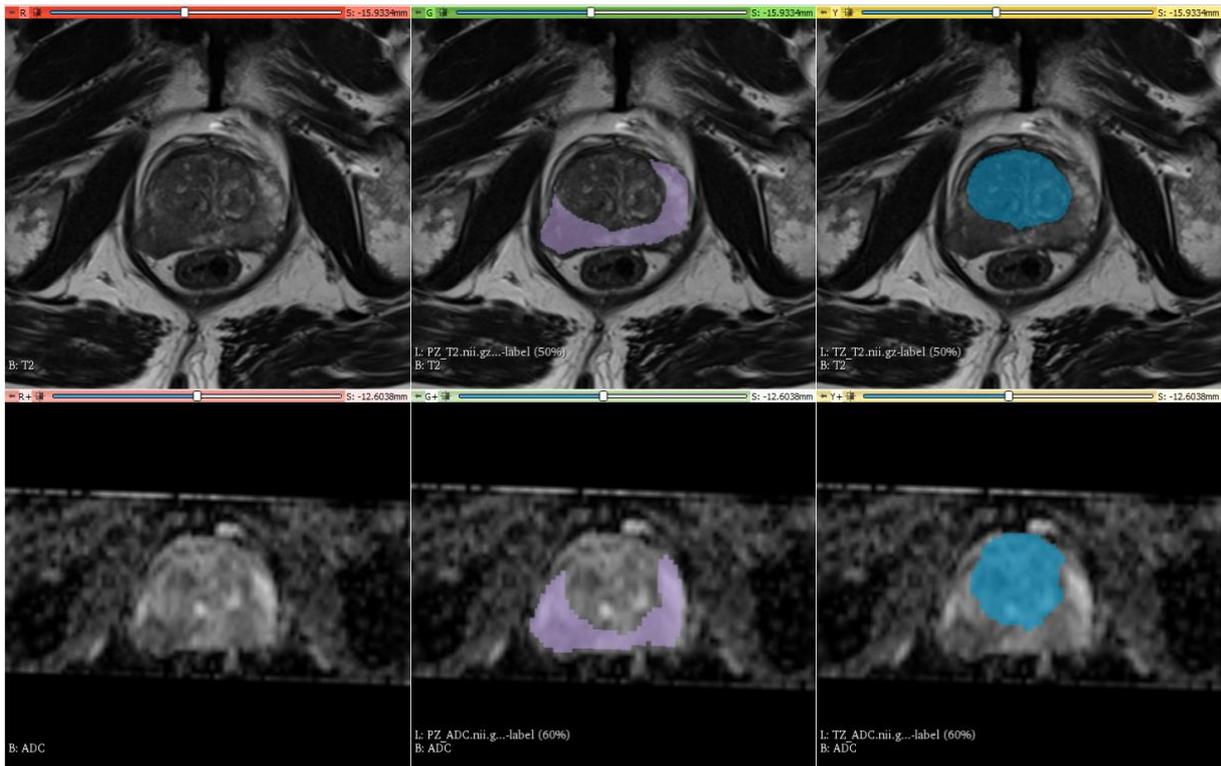


Abb. 10: Manuelle Segmentierung eines repräsentativen Falls (Gleason 4+5, PI-RADS 5, Läsion in PZ). Gezeigt sind die Delineationen der PZ (oben mittig) und TZ (oben rechts) in  $T_2w$ , sowie der PZ (unten mittig) und TZ (unten rechts) in der ADC-map (unten rechts).

Die ROIs wurden im Anschluss durch eine Radiologin oder einen Radiologen revidiert und ggf. korrigiert.

Diese Segmentierungen konnten nun als Ausgangspunkt zum Training der Machine-Learning Modelle verwendet werden.

### 3.5. Preprocessing

Das digitale Preprocessing der Daten dieser Studie wurde mit den Funktionen der open-source Plattform PyRadiomics v3.0 durchgeführt [277]. Diese basieren ihrerseits auf den Funktionen von SimpleITK [314, 315].

Die  $T_2w$ -Bilder wurden auf Studienbasis normalisiert. Die ADC-Bilder wurden über den gesamten Datensatz verteilt normalisiert, um die ursprünglichen Verteilungen der ADC-Werte beizubehalten.

Es wurden die Filter Wavelet und Laplacian of Gaussian mit den Sigma-Werten 5 mm und 7 mm auf die Daten angewandt.

Um die Performance des Modells zu verbessern, wurde die Feature-Engineering Technik „Binning“ oder auch „Diskretisierung“ angewendet. Dabei werden Datensätze in Intervalle („bins“, engl. Für „Behälter“) mit bestimmten Attributen (in dieser Studie First-Order Statistics) eingeordnet, deren optimale Wertespannbreite („Bin width“) man entsprechend der Anzahl der Daten festlegt. Die Datensätze in einem Bin werden dann mit einem repräsentativen Wert subsummiert „interval label“, welche dann zur weiteren Analyse herangezogen werden. Dies reduziert als Quantisierungsmethode die Anzahl der auszuwertenden Daten und kann

durch Glättung von Abweichungen in den Daten insbesondere bei kleinen Datensätzen dazu beitragen, ein Overfitting zu vermeiden. In dieser Studie wurden 32 Bins gebildet.

### 3.6. Datenaugmentierung

In dieser Studie wurden zum Balancing der unausgewogenen Datensätze bzw. zum Augmentieren der Minderheitsklasse der klinisch nicht signifikanten PCa neue Daten im Sinne eines Up-Sampling bzw. Over-Sampling (Erhöhung der Minderheitsklasse) generiert. Zur Erzeugung synthetischer Daten für die Minderheitsklasse wurde SMOTE verwendet [281]. Des Weiteren wurden für den Deep Learning-Ansatz zur automatisierten Segmentierung und Klassifikation die Preprocessing-Techniken Random Flip, Random Rotation in Range ( $[-\pi/12, \pi/12]$ ) und Random Zoom (Range 0.9x - 1.1x) mit einer Wahrscheinlichkeit von 0,5 angewandt.

### 3.7. Radiomics Features - Extraktion und Selektion

Mit dem open-source Python Paket PyRadiomics v3.0 wurden Radiomics Features extrahiert [277]. Der Feature-Extraktion als Basis dienten die durch manuelle Segmentierung von den T<sub>2w</sub>- und ADC-Bildern abgeleiteten ROIs.

Um ein Subset der prädiktivsten und am wenigsten korrelierten Features auszuwählen, wurde anschließend der mRMR-Algorithmus angewandt [316]. Mit dieser Methode wurden 10 Features selektiert. Die Feature-Extraktion und Selektion entsprechen den Feature-Definitionen, die von der Imaging Biomarker Standardization Initiative vorgegeben werden [317].

### 3.8. Extrahierte Radiomics Features

Insgesamt wurden 86 Features extrahiert. Die extrahierten Features lassen sich den Feature-Klassen First-order-statistics, Texture Features (Gray-Level Run Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), Gray-Level Difference Matrix (GLDM)) und Shape-Features zuordnen.

In Tab. 4 werden die Feature-Klassen näher charakterisiert.

Tab. 4: Detaillierte Charakterisierung der einzelnen Feature-Klassen.

| Feature-Klasse       | Features   |
|----------------------|--|
| First-Order Features | Energy, Total Energy, Entropy, Minimum, 10th Percentile, 90th Percentile, Maximum, Mean, Median, Interquartile Range, Range, Mean Absolute Deviation, Robust Mean Absolute Deviation, Root Mean Squared, Standard Deviation, Skewness, Kurtosis, Variance, and Uniformity  |
| Shape Features (3D)  | Mesh Volume, Voxel Volume, Surface Area, Surface Area to Volume Ratio, Sphericity Compactness, Spherical Disproportion, Maximum 3D Diameter, Maximum 2D Diameter (Slice), Maximum 2D Diameter, Maximum 2D Diameter, Major Axis Length, Minor Axis Length, Least Axis Length, Elongation, and Flatness  |
| GLRLM                | Short-Run Emphasis, Long-Run Emphasis, Gray-Level Nonuniformity, Gray-Level Nonuniformity Normalized, Run Length Nonuniformity, Run Length Nonuniformity Normalized, Run Percentage, Gray-Level Variance, Run Variance, Run Entropy, Low Gray-Level Run Emphasis, High Gray-Level Run Emphasis, Short-Run Low Gray-Level Emphasis, Short-Run High Gray-Level Emphasis, Long-Run Low Gray-Level Emphasis, and Long-Run High Gray-Level Emphasis             |
| GLSZM                | Small-Area Emphasis, Large-Area Emphasis, Gray-Level Nonuniformity, Gray-Level Nonuniformity Normalized, Size-Zone Nonuniformity, Size-Zone Nonuniformity Normalized, Zone Percentage, Gray-Level Variance, Zone Variance, Zone Entropy, Low Gray-Level Zone Emphasis, High Gray-Level Zone Emphasis, Small-Area Low Gray-Level Emphasis, Small-Area High Gray-Level Emphasis, Large-Area Low Gray-Level Emphasis, and Large-Area High Gray-Level Emphasis |
| GLDM                 | Small Dependence Emphasis, Large Dependence Emphasis, Gray-Level Nonuniformity, Dependence Nonuniformity, Dependence Nonuniformity Normalized, Gray-Level Variance, Dependence Variance, Dependence Entropy, Low Gray-Level Emphasis, High Gray-Level Emphasis, Small Dependence Low Gray-Level Emphasis, Small Dependence High Gray-Level Emphasis, Large Dependence Low Gray-Level Emphasis, and Large Dependence High Gray-Level Emphasis               |

### 3.9. Training-Validierungs-Split

Die Kohorte wurde in einen Trainings-Datensatz (80%,  $n = 238$ ) und einen Validierungs-Datensatz (20%,  $n = 59$ ) Beispielen aufgeteilt.

### 3.10. Klassifikationsmodelle: Selektion und Validierung

Die vier Klassifikator-Modelle Logistische Regression, Linear Support Vector Machine, Random Forest und XGBoost wurden mithilfe des Programms scikit-learn angewendet [318]. Jedes dieser vier Modelle nahm Klassifikationen basierend auf den extrahierten Radiomics Features vor. Im Anschluss erfolgte die Selektion des besten Modells. Dazu wurden die Modelle mit 5-fach-Kreuzvalidierung („5-fold cross-validation“, CV) und Bootstrap Bias-Korrektur (BBC-CV) evaluiert und verglichen.

### 3.11. Performance-Maße

Die statistische Auswertung erfolgte mit R Studio [319].

Alle Daten sind als Mittelwerte  $\pm$  Standardabweichung angegeben.

Alle statistischen Tests waren 2-seitig, und ab einem  $p$ -Wert  $< 0.05$  wurde von einer statistischen Signifikanz ausgegangen.

Die kontinuierlichen Variablen wurden auf Normalverteilung geprüft.

Mit dem  $t$ -Test für unabhängige Stichproben wurden die errechneten Mittelwerte verglichen.

Die Ergebnisse wurden in einer Confusion Matrix dargestellt. Es wurde die (mittlere) Accuracy der Modelle berechnet.

Im Anschluss wurde als Qualitätsmaß zur Evaluation und Visualisierung der Klassifikator-Performances eine Grenzwertoptimierungskurve (AUC) aufgetragen.

### 3.12. Alternativer Ansatz: Deep Learning

Außerdem wurde ein Deep Learning-Ansatz durchgeführt. Abb. 11 zeigt allgemeine Arbeitsschritte zur Entwicklung eines Deep-Learning Algorithmus zur Diagnose von PCa, (wobei im Rahmen dieser Studie keine klinischen Daten, sondern nur Bilddaten als Eingang verwendet wurden).

Hierzu eignen sich verschiedene „pre-trained models“. Diese sind in Online-Plattformen wie dem in dieser Studie verwendeten Medical Open Network for Artificial Intelligence („MONAI“) frei verfügbar und beinhalten bereits erlernte Gewichte, sodass sie sofort zur Segmentierung der Prostata angewendet werden können [320]. In dieser Studie wurde für den Deep Learning-Ansatz das CNN DenseNet-121 verwendet [321].

Durch Concatenation („Verkettung“) wurden die Matrices verschiedener Datentypen kombiniert und in eine einzige Matrix integriert. Es wurden auf eine Maskenregion

zugeschnittene und verkettete T<sub>2</sub>w-Sequenzen und ADC-maps entsprechend einer von Liu et al. beschriebenen Methode als Input für das prädiktive Deep Learning-Modell verwendet [322].

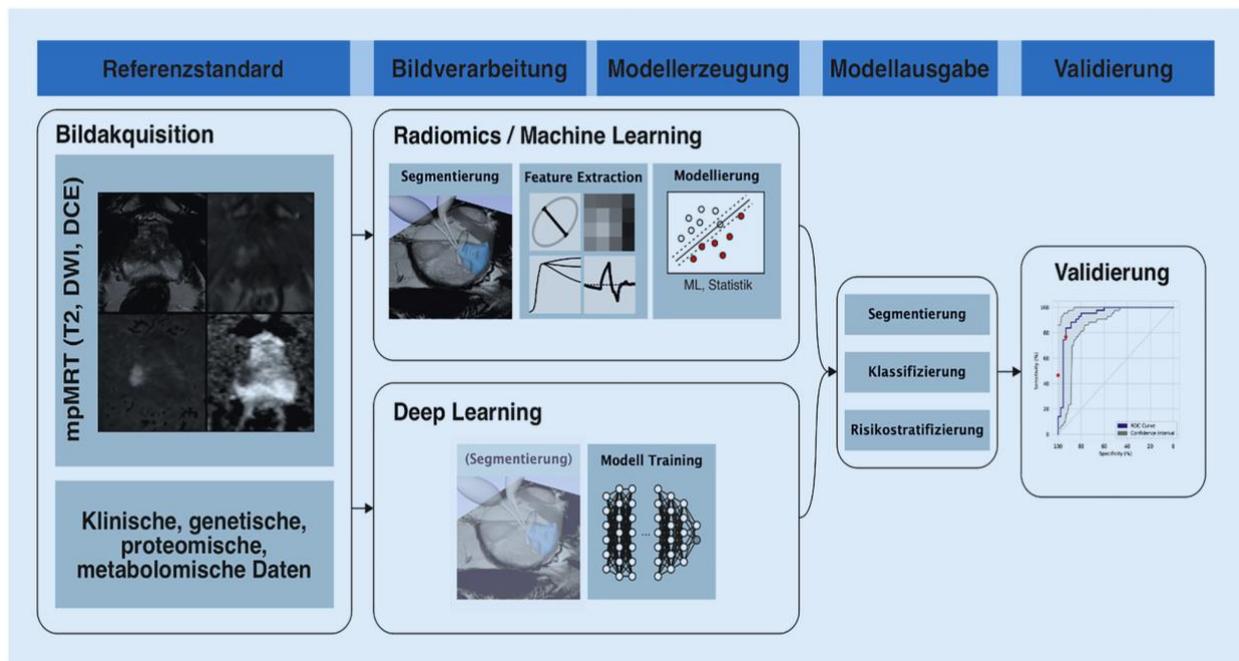


Abb. 11: „Darstellung der Arbeitsschritte zur Erzeugung eines Deep Learning-computergestützten Diagnostiksystems. Referenzstandard (links): basierend auf mpMRT der Prostata und klinischen Daten. Bildverarbeitung/Modellerzeugung (Mitte): Mit Radiomics werden anhand bekannter Features Vorhersagemodelle erzeugt, beim Deep Learning werden diese Features selbstständig vom System bestimmt. Modellausgabe (rechts): Die mit manuell selektierten Radiomics oder mit Deep Learning erstellten Modelle werden zur Segmentierung, Klassifizierung und Risikostratifizierung angewandt. Validierung (ganz rechts): Die Modelle werden überprüft“ [133].

Anmerkung: im Rahmen dieser Studie wurden keine klinischen Daten, sondern nur Bilddaten als Eingang verwendet.

Mit freundlicher Genehmigung von: Springer Nature / Der Radiologe 60, 48-55 [Artificial intelligence and radiomics in MRI-based prostate diagnostics]. Hamm, C.A., Beetz, N.L., Savic, L.J., and Penzkofer, T. (2020).

### 3.13. Sequence-Model

Diese Studie verfolgte einen schichtbasierten Ansatz. Ein Vergleich mit dem PI-RADS Klassifikationssystem, welches Läsionen im gesamten Organ betrachtet, ist mit diesem schichtbasierten Ansatz nicht sinnvoll. Jedoch ist dies vor dem Hintergrund des Ziels, die Bilddaten der gesamten Prostata als Input eingeben und ohne vorherige Läsionssegmentierung durch einen Klassifikator untersuchen zu können, für die klinische Anwendung interessant. Um diese Vergleichbarkeit zu ermöglichen, wurde der Deep Learning-Ansatz durch ein sogenanntes Sequence-Model ergänzt [323]. Das Sequence-Model ist ein Verfahren, das sich in einem von der Radiological Society of North America (RSNA) und der Society of Thoracic Radiology (STR) ausgerichteten großen Machine Learning-Wettbewerb zur Detektion und Klassifikation von pulmonalen Embolien als äußerst erfolgreich erwies und den ersten Platz belegte.

Mit dem Sequence-Model bleibt der grundsätzliche 2D- bzw. schichtbasierte Ansatz erhalten, jedoch werden Deep-Learning Features aus jeder der einzelnen Schichten durch das Modell übereinandergelegt, sodass eine hochdimensionale 3D-Matrix entsteht. Ähnlich einer Radiologin bzw. eines Radiologen wird jede Schicht einzeln zur Meinungsbildung betrachtet, diese „Meinung“ im Kontext mit jeder nächsten Schicht neu angepasst, und integrativ zu einer Diagnose verarbeitet. Somit wird sekundär ein „study-based approach“, also ein die gesamte MRT-Studie abbildender Ansatz umgesetzt. Diese Lösung ermöglichte in dieser Arbeit einen Vergleich mit PI-RADS.

## 4. ERGEBNISSE

### 4.1. Klinische Charakteristika der Patientenkohorte

Tab. 5 stellt die klinischen Charakteristika der Patientenkohorte dieser Studie dar.

Tab. 5: Klinische Charakteristika der Patientenkohorte.

|   | <b>Training</b> | <b>Validierung</b> |
|---|-----------------|--------------------|
| <b>Patienten insgesamt</b>  | 238             | 59                 |
| • Benigne Läsion/en   | 77              | 23                 |
| • Indexläsionen ISUP Grad 1 und 2 (klinisch nicht signifikantes PCa)              | 105             | 10                 |
| • Indexläsionen ISUP Grad $\geq 3$ (GS $\geq 4 + 3$ , klinisch signifikantes PCa) | 56              | 26                 |
| <b>Davon:</b>   |                 |                    |
| • ISUP Grad 1 (GS 3 + 3)  | 53              | 10                 |
| • ISUP Grad 2 (GS 3 + 4)  | 52              | 13                 |
| • ISUP Grad 3 (GS 4 + 3)  | 37              | 7                  |
| • ISUP Grad 4 (GS 4 + 4 / 3 + 5 / 5 + 3)  | 12              | 5                  |
| • ISUP Grad 5 (GS 4 + 5 / 5 + 4 / 5 + 5)  | 7               | 1                  |
| <b>Läsionen insgesamt</b>   | 441             | 189                |
| <b>Davon:</b>   |                 |                    |
| • benigne   | 169             | 82                 |
| • klinisch nicht-signifikante PCa   | 88              | 20                 |
| • klinisch signifikante PCa   | 184             | 87                 |
| <b>DRU</b>  |                 |                    |
| • Positiv   | 58              | 13                 |
| • Negativ   | 180             | 46                 |
| <b>Annotierte Schichten insgesamt</b>   | 2025            | 544                |
| <b>Mittlerer PSA-Wert</b>   | 11,9 $\pm$ 13,2 | 10,6 11,1          |
| <b>Mittleres Prostatavolumen</b>  | 55,2 $\pm$ 29,0 | 57,0 $\pm$ 32,1    |

Abb. 12 gibt die Verteilung der Kategorien „benigne Läsion/en“, „Indexläsionen ISUP Grad 1 und 2“ und „Indexläsionen ISUP Grad  $\geq 3$ “ in der Kohorte graphisch wieder.

● benigne Läsionen   ● ISUP Grad 1 und 2   ● ISUP Grad  $\geq 3$

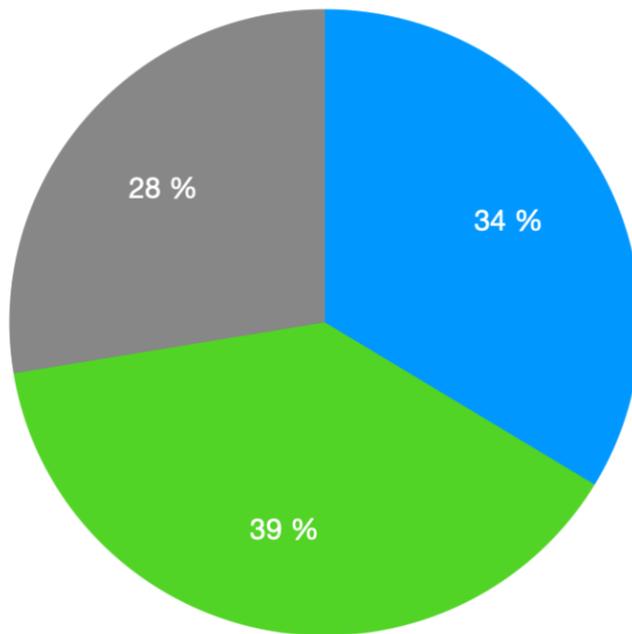


Abb. 12: Verteilung „benigne Läsionen“, „ISUP Grad 1 und 2“ und „ISUP Grad  $\geq 3$ “ in der Patientenkohorte.

#### 4.2. Selektion der besten Features für den ML-Ansatz: Ausgewählte Features

Die Features wurden sowohl aus den T<sub>2</sub>-Aufnahmen als auch den ADC-maps extrahiert. Die im Anschluss mit dem mRMR-Algorithmus selektierten Features sind die Folgenden:

- 'original\_shape\_MinorAxisLength\_t2',
- 'original\_firstorder\_Minimum\_adc',
- 'wavelet-LHL\_firstorder\_Mean\_adc',
- 'original\_shape\_Flatness\_adc',
- 'wavelet-LLL\_firstorder\_Skewness\_adc',
- 'wavelet-LLL\_firstorder\_TotalEnergy\_t2',
- 'wavelet-HHL\_firstorder\_Mean\_t2',
- 'original\_shape\_Elongation\_t2',
- 'wavelet-HLL\_firstorder\_Skewness\_adc',
- 'original\_shape\_Sphericity\_t2']

Die in dieser Studie verwendeten Feature Klassen lassen sich dementsprechend Shape-Features, First-Order-Features und Textur-Features zuordnen.

#### 4.3. Selektion des besten ML-Modells mit 5-facher Kreuzvalidierung

Tab. 6 stellt die mittlere (Diagnose-) Genauigkeit als hier verwendete Performance-Metrik der Klassifikationen der verschiedenen Modelle dar. Das über PyRadiomics durchgeführte „Shape-based“ Verfahren basierte auf zwei von der Plattform angebotenen Input-Varianten: Beim „Single slice input“ wird eine einzelne Schicht herangezogen. Beim „Three slice input“ hingegen werden die beiden direkt benachbarten Schichten herangezogen. Dies gibt der Analyse mehr Kontext durch Nachbarstrukturen, die auf den beiden anliegenden Schichten erfasst sind. Es wurden zum Training die Segmentierungen „Organ gesamt“ (2025 annotierte Schichten) verwendet.

Tab. 6: Mittlere Diagnosegenauigkeit der einzelnen Klassifikationsmodelle mit Single-slice- und Three-slice-Input. Mean Accuracy = mittlere Genauigkeit.

| Input        | ROI              | Klassifikationsmodell  | Mean Accuracy        |
|--------------|------------------|------------------------|----------------------|
| Single slice | Gesamte Prostata | Logistische Regression | 0.61 +/- 0.05        |
|              |                  | SVM                    | 0.61 +/- 0.06        |
|              |                  | Random Forest          | 0.59 +/- 0.06        |
|              |                  | XGBoost                | 0.60 +/- 0.05        |
| Three slices |                  | Logistische Regression | 0.61 +/- 0.09        |
|              |                  | <b>SVM</b>             | <b>0.63 +/- 0.09</b> |
|              |                  | Random Forest          | 0.60 +/- 0.07        |
|              |                  | XGBoost                | 0.62 +/- 0.04        |

Es wird ersichtlich, dass die SVM mit einem Three-slice-Input am Trainings-Datensatz das beste Ergebnis erreicht (mittlere Diagnosegenauigkeit von 0,63 +/- 0,09). Diese Ergebnisse wurden mit der Einstellung folgender Parameter erzielt:

- Anzahl der Features: 10
- Feature-Klassen: alle aus PyRadiomics (Shape-, Textur- und First-Order-Features)
- Features extrahiert aus: T<sub>2</sub>w und ADC
- Angewandte Filter: kein Filter (original), Laplacian of Gaussian mit 3 Sigma (5 mm, 7 mm und Wavelet)

#### 4.4. Ergebnisse des SVM-Klassifikators mit Three-slice-Input

Die Performance des SVM-Modells mit Three-slice-Input im Trainings-Datensatz wurde mit 5-facher Kreuzvalidierung evaluiert. Tab. 7 stellt die Performance bei der Klassifikation in die drei Kategorien „kein PCa“, „non-csPCa“, und „csPCa“ im Trainings-Datensatz in einer Confusion Matrix dar.

Tab. 7: Ergebnisse des SVM-Klassifikationsmodells am Trainings-Datensatz, dargestellt in einer Confusion Matrix.

|           | Als benigne klassifiziert | Als non-csPCa klassifiziert | Als csPCa klassifiziert |
|-----------|---------------------------|-----------------------------|-------------------------|
| Benigne   | 127                       | 9                           | 33                      |
| Non-csPCa | 18                        | 8                           | 62                      |
| csPCa     | 20                        | 20                          | 144                     |

Der Trainings-Datensatz enthielt insgesamt 169 benigne Läsionen, 88 non-csPCa, und 184 csPCa. Die AUC betrug 0,88 mit einer falsch-positiven Fehlklassifikation von insgesamt 42 tatsächlich benignen Läsionen als maligne (9 benigne Läsionen wurden als non-csPCa und 33 benigne Läsionen als csPCa eingestuft). 127 benigne Läsionen wurden korrekt als solche klassifiziert. Falsch-negativ wurden 20 csPCa als benigne klassifiziert. Von 184 csPCa wurden 20 als non-csPCa fehlklassifiziert. Korrekt als csPCa wurden 144 Läsionen klassifiziert. Abb. 13 zeigt die Ergebnisse der 5-fachen Kreuzvalidierung der SVM-Klassifikationen als ROC. Die Klassifikation in „benigne Läsion“ oder „Vorliegen eines PCa“ (sowohl non-cs PCa als auch csPCa) wurde mit einer AUC von 0,88 durchgeführt.

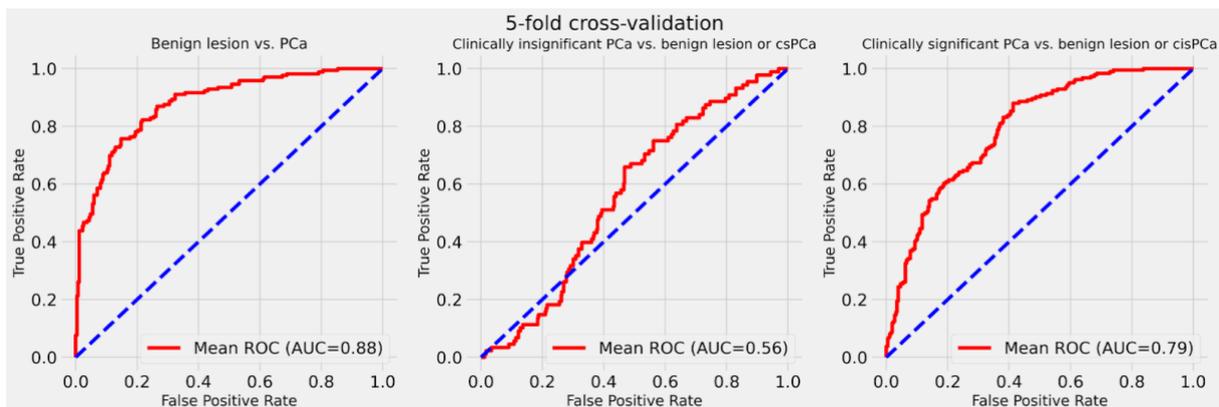


Abb. 13: ROC der SVM-Klassifikationen am Trainings-Datensatz. y-Achse: Richtig-positiv-Rate, x-Achse: Falsch-positiv-Rate. Links: „benigne Läsion“ vs. „PCa“. Mitte: „non-csPCa“ (clinically insignificant PCa) vs. „benigne Läsion oder csPCa“. Rechts: „csPCa“ vs. „benigne Läsion oder non-csPCa“.

#### 4.5. Ergebnisse des SVM-Klassifikators mit Three-slice-Input am Validierungs-Datensatz

Zur Evaluation von Performance und Generalisierbarkeit des Klassifikationsmodells wurde es auf einen separaten Validierungs-Datensatz angewendet, der zuvor noch nicht zur Modellentwicklung verwendet wurde. Tab. 8 stellt die Ergebnisse des SVM-Klassifikators am Validierungs-Datensatz (544 annotierte Schichten, darin insgesamt 82 benigne Läsionen, 20 non-csPCa, und 87 csPCa) in einer Confusion Matrix dar.

Tab. 8: Ergebnisse der Anwendung des SVM-Klassifikationsmodells am Validierungs-Datensatz, dargestellt als Confusion Matrix.

|           | Als benigne klassifiziert | Als non-csPCa klassifiziert | Als csPCa klassifiziert |
|-----------|---------------------------|-----------------------------|-------------------------|
| Benigne   | 73                        | 3                           | 6                       |
| Non-csPCa | 7                         | 2                           | 11                      |
| csPCa     | 22                        | 0                           | 65                      |

Abb. 14 zeigt die Performance des Modells am Validierungs-Set als ROC.

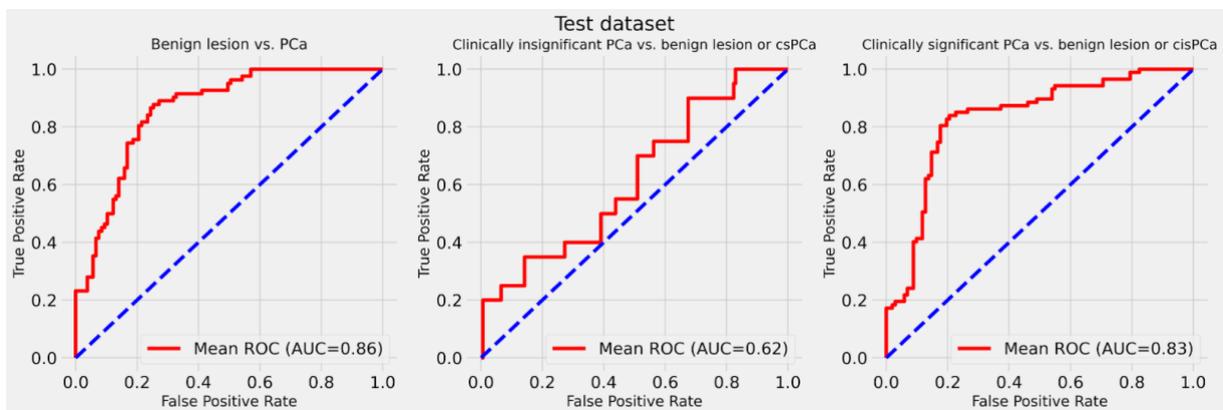


Abb. 14: ROC der SVM-Klassifikationen am Validierungs-Datensatz. y-Achse: Richtig-positiv-Rate, x-Achse: Falsch-positiv-Rate. Links: „benigne Läsion“ vs. „PCa“. Mitte: „non-csPCa“ (clinically insignificant PCa) vs. „benigne Läsion oder csPCa“. Rechts: „csPCa“ vs. „benigne Läsion oder non-csPCa“.

Die Klassifikation des SVM-Modells in „benigne Läsion“ oder „PCa“ (sowohl non-csPCa als auch csPCa) wurde mit einer AUC von 0,86 durchgeführt. Klinisch nicht signifikante PCa wurden von benignen Läsionen und signifikanten PCa mit einer AUC von 0,62 unterschieden. Klinisch signifikante PCa konnten von benignen Läsionen und insignifikanten PCa mit einer AUC von 0,83 unterschieden werden.

#### 4.6.Zonaler Ansatz: Ergebnisse

Ergänzend zum organbasierten Ansatz wurde die getrennte Analyse der peripheren Zone (ROI PZ) und der Transitionalzone (ROI TZ) durchgeführt. Abb. 15 veranschaulicht die Performance des SVM-Modells für die periphere Zone, dargestellt als ROC. Es wurde bei der Klassifikation von benigner Läsion oder PCa eine AUC von 0,75 erreicht.

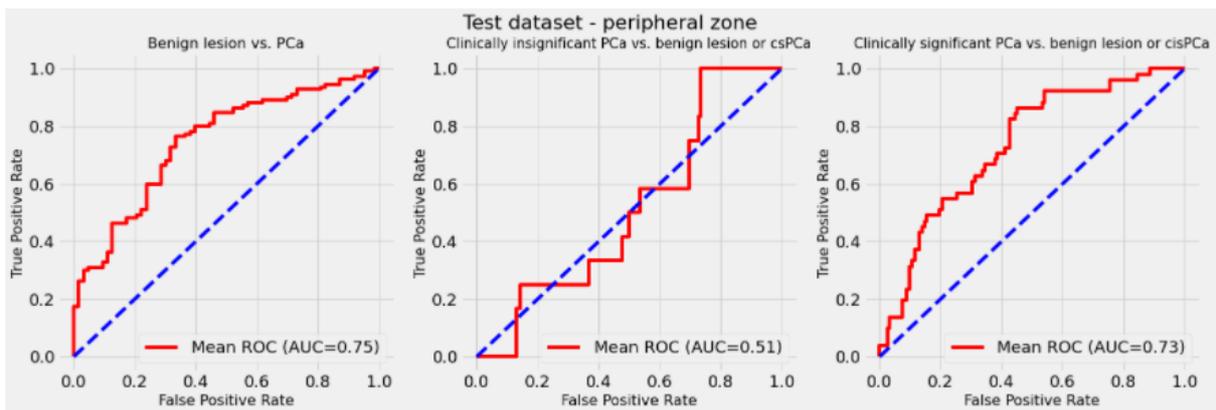


Abb. 15: ROC der SVM-Klassifikationen ausschließlich für die periphere Zone (peripheral zone). Y-Achse: Richtig-positiv-Rate, x-Achse: Falsch-positiv-Rate. Links: „benigne Läsion“ vs. „PCa“. Mitte: „non-csPCa (clinically insignificant PCa)“ vs. „benigne Läsion oder csPCa“. Rechts: „csPCa“ vs. „benigne Läsion oder non-csPCa“.

Abb. 16 zeigt die ROC für die Klassifikation der Transitionalzone. Hier wurde eine AUC von 0,64 für die Einschätzung in benigne Läsion oder PCa erreicht.

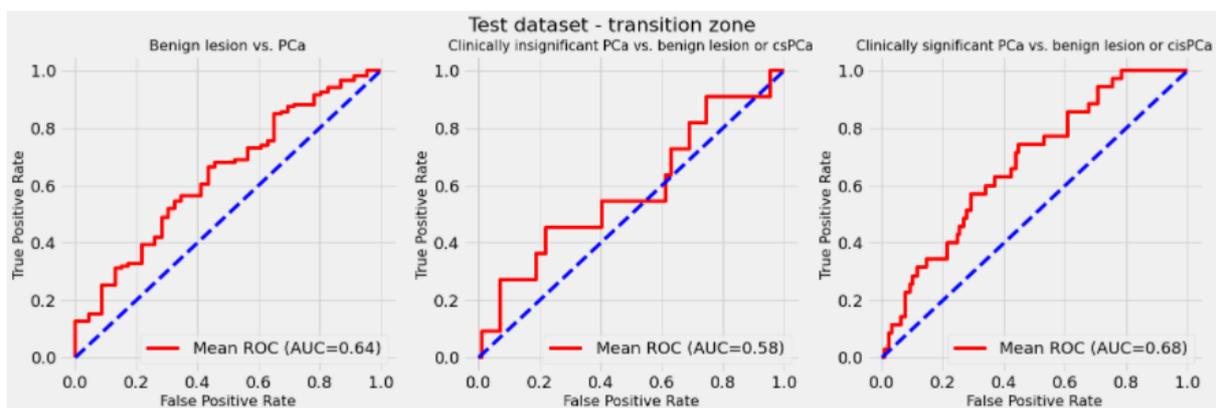


Abb. 16: ROC der SVM-Klassifikationen ausschließlich für die Transitionalzone (transition zone). Y-Achse: Richtig-positiv-Rate, x-Achse: Falsch-positiv-Rate. Links: „benigne Läsion“ vs. „PCa“. Mitte: „non-csPCa“ (clinically insignificant PCa) vs. „benigne Läsion oder csPCa“. Rechts: „csPCa“ vs. „benigne Läsion oder non-csPCa“.

#### 4.7. Deep Learning-Ansatz: Ergebnisse

Die Klassifikationen des DenseNet121-Modells ergab eine mittlere Diagnosegenauigkeit von 0,59 für sowohl den Single-slice- als auch den Three-slice-Input. Tab. 9 zeigt die Ergebnisse für den Deep Learning-Ansatz.

*Tab. 9: Ergebnisse der 5-fachen Kreuzvalidierung der Klassifikationen durch das Deep Learning-Modell. Mean Accuracy = Mittlere Genauigkeit.*

| Input        | ROI             | Klassifikationsverfahren    | Mean Accuracy |
|--------------|-----------------|-----------------------------|---------------|
| Single slice | Prostata gesamt | Deep Learning (DenseNet121) | 0.59 +/- 0.04 |
| Three slices |                 |                             | 0.59 +/- 0.05 |

#### 4.8. Sequence-Model-Ansatz: Ergebnisse

Das Sequence-Model zur studienbasierten Klassifikation zum Vergleich mit der Einordnung in „benigne“ bzw. „PCa“ anhand des durch einen Reader vergebenen PI-RADS Scores erfolgte an einem Datensatz bestehend aus 33 Patientendatensätzen. Diese beinhalteten 19 benigne Läsionen, 6 non-csPCa, und 18 csPCa. Es wurden ebenfalls die „Prostata gesamt“-ROI verwendet.

Als Input wurden die vom DenseNet21-Modell extrahierten Features verwendet.

Die AUC für die Klassifikation des Sequence-Modells in „benigne Läsion“ oder „PCa“ betrug 0,81. Die AUC für die Klassifikation unter Anwendung der PI-RADS-Klassifikation durch den menschlichen Reader lag für diesen Datensatz bei 0,77 (vgl. Abb. 17).

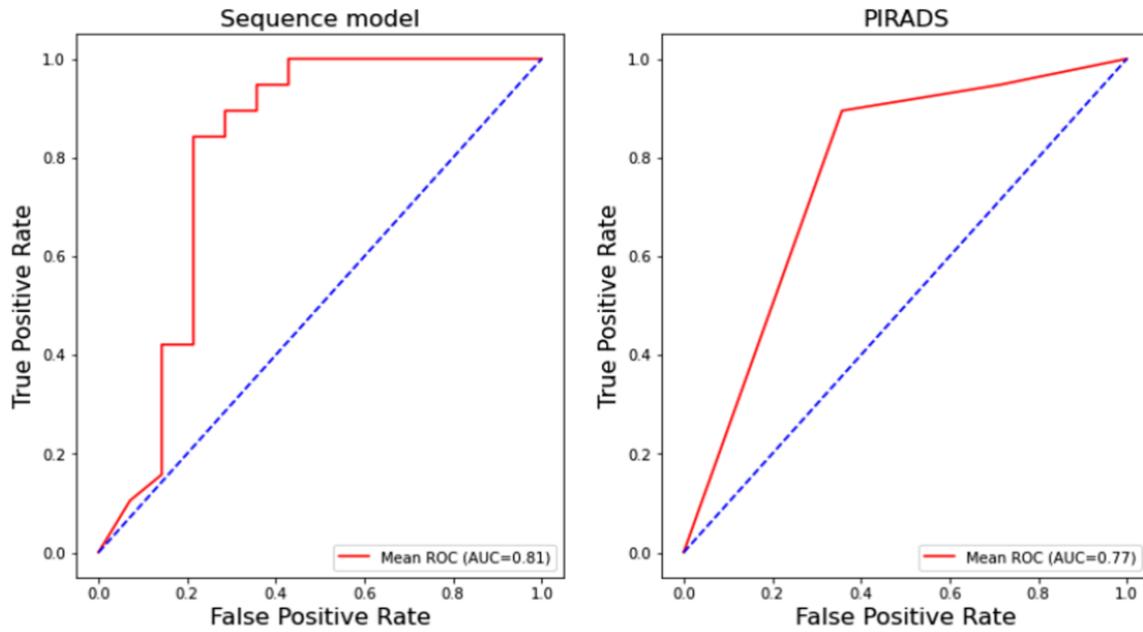


Abb. 17: AUC für die Klassifikation in „benigne Läsion“ oder „PCa“ durch das Sequence-Model (links) und mittels PI-RADS-Klassifikationssystem (rechts).

Die Klassifikation in „cs-PCa“ oder „non-csPCa oder benigne Läsion“ wurde mit einer AUC von 0,86 durch das Sequence-Model und einer AUC von 0,83 mittels PI-RADS-Klassifikation durchgeführt (vgl. Abb. 18).

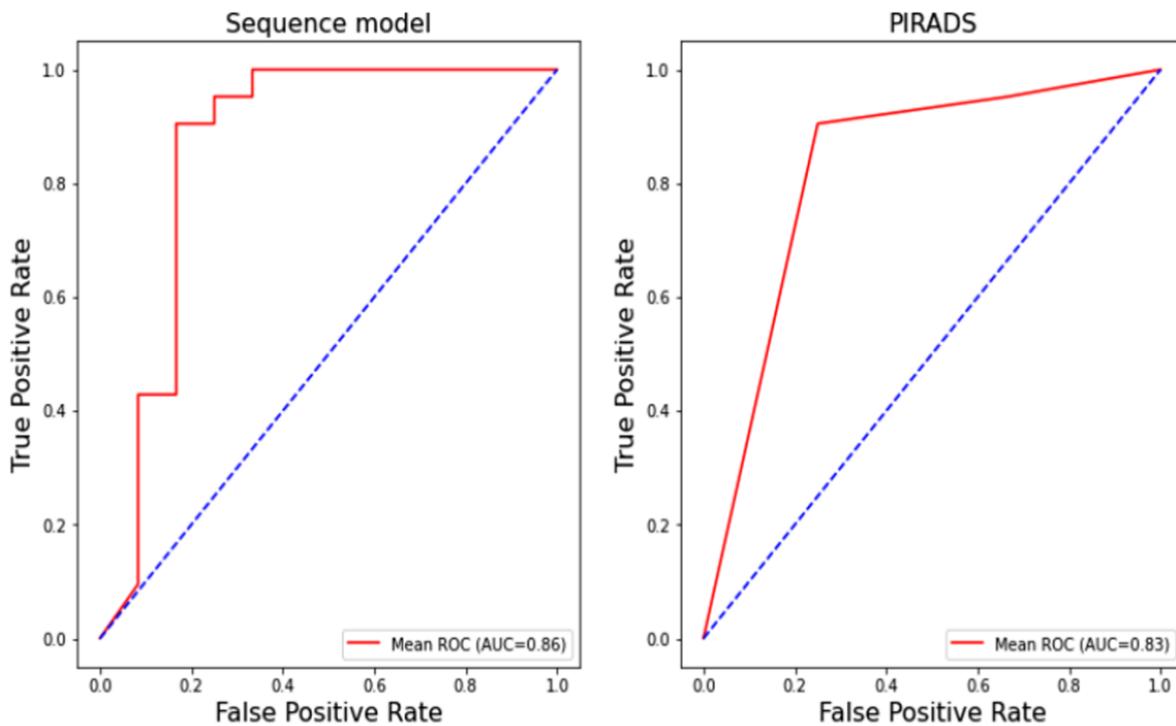


Abb. 18: AUC für die Klassifikation in „csPCa“ oder „non-csPCa oder benigne Läsion“. Links: Klassifikation des Sequence-Models. Rechts: Klassifikation mit PI-RADS.

## 5. DISKUSSION

### 5.1. Diskussion der Ergebnisse dieser Studie

In dieser Arbeit sollte untersucht werden, ob ein Machine Learning-Modell, basierend auf aus mpMRT der Prostata extrahierten Radiomics Features, eine erfolgreiche Differenzierung zwischen klinisch signifikanten und klinisch nicht signifikanten PCa durchführen kann. Hierzu wurde im Gegensatz zu einem Läsions-basierten Ansatz ein Schicht-basierter Ansatz verfolgt. Außerdem wurde ein klassischer Radiomics-Ansatz mit einem Deep Learning-Ansatz verglichen.

Das übergeordnete Ergebnis dieser Studie ist, dass die Verarbeitung von aus mpMRT extrahierten Radiomics Features Vorhersagen zur klinischen Signifikanz von PCa zulässt, die in ihrer Aussagekraft mit derjenigen von konventionellen Befundungsmethoden wie PI-RADS vergleichbar sind.

Im Detail: Der schichtbasierte Radiomics-Ansatz erbrachte am Validierungs-Datensatz bei der Läsions-Klassifikation in „benigne“ oder „PCa“ mit dem SVM-Klassifikator eine hohe AUC von 0,88. Dies ist als ein gutes Ergebnis zu werten (vgl. Tab. 10 zur Interpretation von AUC-Scores).

Bei der Anwendung der Klassifikatoren auf die 10 selektierten Features (den Klassen First-Order, Shape, Textur zugehörig) erzielte der SVM-Klassifikator die besten Ergebnisse (mittlere Genauigkeit von 0,63 +/- 0,09). Am Validierungs-Datensatz erzielte das SVM-Modell eine ebenfalls hohe AUC von 0,86.

Die Unterscheidung zwischen „klinisch signifikantes PCa“ und „benigne Läsion“ bzw. „klinisch nicht signifikantes PCa“ wurde mit mäßigem bis gutem Ergebnis durchgeführt (AUC von 0,79 am Trainings-Datensatz und AUC von 0,83 am Validierungs-Datensatz).

Die zonalen Ergebnisse fielen schlechter aus als die des organbasierten Ansatzes (AUC von 0,75 für die periphere Zone, und 0,64 für die Transitionalzone).

Die Ergebnisse dieser Studie entsprechen weitgehend der aktuellen Studienlage. Eine Studie an 139 Patienten zur Differenzierung zwischen klinisch signifikanten und nicht signifikanten PCa kam ebenfalls zu dem Ergebnis, dass eine mpMRT-basierte Radiomics-Signatur (aus 9 Radiomics Features) Potential als nicht-invasive Methode zur PCa-Diagnostik hat [324].

Eine weitere Studie an 240 Patienten untersuchte, ob ein T<sub>2w</sub>- Radiomics-basiertes ML-Modell klinisch signifikante PCa unter PI-RADS 3-Läsionen identifizieren kann. Dieses erreichte eine AUC von 0,76 [325].

Auch eine weitere Studie mit dem Ziel, klinisch signifikante PCa unter Läsionen der PI-RADS Kategorie 3 zu identifizieren, kam zu dem Schluss, dass ein ML-Modell durchaus zur Prädiktion signifikanter PCa in der Lage ist und zur Entscheidungsfindung vor Biopsien beitragen kann [326].

Eine ebenfalls 10 Features integrierende Radiomics-Signatur, extrahiert aus T<sub>2w</sub>, DWI und ADC-Bildern erreichte ein sehr gutes Ergebnis bei der Differenzierung zwischen signifikanten und indolenten PCa (AUC von 0,95 am Trainings-Datensatz, 0,86 in der internen Validierung, 0,81 in der externen Validierung) [327].

Eine andere Studie verglich die Performance eines auf biparametrischer MRT basierten Radiomics-ML-Modells mit einem kombinierten Radiomics- und klinisch basierten Modell, sowie mit einem allein klinisch basierten Modell (Risikofaktoren: Alter, Drüsenvolumen, PSA-Wert usw.). Die höchste prädiktive Genauigkeit erbrachten das Radiomics- und das kombinierte Modell (AUC von jeweils 0,98; klinisch basiert: AUC von 0,79), sodass abgeleitet wurde, dass

die Diagnosegenauigkeit bei der Detektion klinisch signifikanter PCa durch die Anwendung eines Radiomics-basierten ML-Modells verbessert werden kann [328].

Ein systematisches Review zur Performance von ML-Methoden zur PCa-Diagnostik kam zu dem Ergebnis, dass ML-Methoden in einigen Studien eine zufriedenstellende Performance zeigen konnten, dass aber eine Implementierung in ein klinisches Setting aufgrund limitierter Anwendbarkeit und fehlender Standards noch nicht durchführbar ist. Es wird für die benötigte künftige Forschung ein robustes Studiendesign empfohlen. Unsere Studie entspricht in vielen Punkten den genannten Empfehlungen. Hierzu zählen das Management fehlender/unvollständiger Daten, die IBSI-konformen Preprocessing-Schritte wie die Anwendung von Filtern, Datennormalisierung, Feature-Extraktion und -Selektion, die Datensatzkonstruktion in Trainings-, Validierungs- und Testdaten, und der Ausgleich von Klassenungleichgewichten [317]. Die interne Validierung mittels Kreuzvalidierung, sowie die externe Validierung mittels eines Testdatensatzes, und ein alternativer DL-Ansatz waren in unserer Studie gegeben, jedoch kein Test auf extern generierten Daten [329].

Der SVM-Klassifikator wurde in einer erfolgreichen Studie verwendet, die mittels eines Radiomics-basierten ML-Algorithmus die diagnostische Performance von PI-RADS (v2) bei klinisch signifikanten PCa zu steigern versuchte [330]. Auch bei einer multizentrischen Studie zur Texturanalyse zur Aggressivität von PCa wurde die SVM erfolgreich eingesetzt [331].

Die schlechtere Performance des zonalen Ansatzes entspricht nicht der aktuellen Studienlage, allerdings fand keine zonenspezifische Datennormalisierung statt. Bisher konnte mehrfach gezeigt werden, dass ein zonaler Ansatz eine gute Diagnosegenauigkeit für PCa erbringen kann [257, 269, 270, 332-335].

Der Deep Learning-Ansatz erwies sich im ersten Schritt nicht als erfolgreich (mittlere Genauigkeit von 0,59). Bei der Anwendung des Sequence-Modells als zweiter Schritt nach Extraktion der Radiomics Features durch das Deep Learning-Modell ließen sich jedoch für die Klassifikation der Kategorien „benigne Läsion“ oder „PCa“ eine AUC von 0,81, und für die Klassifikation in „klinisch signifikantes PCa“ oder „nicht signifikantes PCa“ bzw. „benigne Läsion“ sogar eine AUC von 0,86 erzielen.

Der Ansatz des Sequence-Modells ermöglichte zudem als hochdimensionales Modell die kombinierte Betrachtung der einzelnen Schichten und legitimierte damit als „studienbasierter“ Ansatz einen Vergleich mit dem klinisch etablierten PI-RADS-Klassifikationssystem. Das Sequence-Modell übertraf die PI-RADS-Performance für die Klassifikation der Kategorien „benigne Läsion“ oder „PCa“ an diesem Test-Datensatz (AUC = 0,81 vs. PI-RADS: AUC = 0,77). Auch bei der Klassifikation der Kategorien „klinisch signifikantes PCa“ oder „nicht signifikantes PCa“ bzw. „benigne Läsion“ fiel die PI-RADS Klassifikation mit einer AUC von 0,83 schlechter aus als die des Sequence-Modells (AUC = 0,86). Der Sequence-Modell-Ansatz stellte sich demnach in dieser Studie auch im Vergleich mit PI-RADS als ein aussagekräftiges diagnostisches Tool zur Prädiktion von PCa dar.

Dass eine mit PI-RADS vergleichbare Performance von ML- oder DL-Verfahren möglich ist, konnten auch einige Studien demonstrieren. So wurde mit der Verwendung von Radiomics aus mpMRT durch ML-Klassifikationssysteme ein höherer positiver Prädiktionswert für PCa erreicht als bei der Anwendung des PI-RADS-Diagnosesystems [239, 330].

In einer weiteren Studie wurde anhand der Bilddaten von 250 Patienten ein DL-Algorithmus entwickelt, der in der Lage war, PCa-Läsionen mit einer Genauigkeit zu detektieren und zu klassifizieren, die mit der des herkömmlichen PI-RADS-Diagnosesystems vergleichbar war [336].

Laut einem bereits angeführten systematischen Review erreichten DL-Ansätze in den meisten Studien eine tendenziell weniger gute Performance, was oft beispielsweise auf die eher kleineren Kohortengrößen zurückführbar ist. DL-Methoden benötigen größere Datenvolumen für stabile Ergebnisse [329, 337].

Andererseits konnte eine Studie beispielsweise ein KNN entwickeln, das mit Erfolg (i.e. Accuracy von 96 % - 98 %) den Gleason Score einer Läsion ausgehend von MRT vorhersagte [338].

Le et al. konnten mit Erfolg ein KNN für zweierlei Aufgaben in der PCa-Diagnostik entwickeln: Zum einen wurde eine Klassifikation in benigne und maligne Läsion durchgeführt, zum anderen wurde zwischen klinisch signifikanten und indolenten PCa unterschieden [339].

Den DL-Methoden gegenüber bietet der Radiomics-Ansatz jedoch mehr Transparenz z.B. in Bezug auf Feature-Selektion [340].

Tab. 10: Interpretation von AUC-Scores [341].

| AUC-Score   | Interpretation |
|-------------|----------------|
| 0,90 – 1    | exzellent      |
| 0,80 – 0,90 | gut            |
| 0,70 – 0,80 | mäßig          |
| 0,60 – 0,70 | schlecht       |
| 0,50 – 0,60 | mangelhaft     |

## 5.2.Diskussion von Methoden und Limitationen dieser Studie

Im Folgenden werden hervorzuhebende Aspekte im Hinblick auf die Methodologie dieser Studie diskutiert.

Die Erstellung der mpMRT-Aufnahmen in dieser Studie erfolgte gemäß den Richtlinien von PI-RADS v2 [342].

Eine größere Kohorte wäre prinzipiell wünschenswert, um bessere Ergebnisse zu erzielen. Außerdem lagen Klassen-Imbalancen innerhalb der Datensätze in Bezug auf die drei Gruppen „benigne Läsion“, „klinisch nicht-signifikantes PCa“, und „klinisch signifikantes PCa“ vor, die als beeinträchtigende Faktoren zu werten sind. Insbesondere in der Gruppe „klinisch nicht-signifikante PCa“ lag verhältnismäßig nur eine geringe Anzahl an Beispielen vor. Für den Radiomics-Ansatz wurden die Klassen-Imbalancen mit SMOTE ausgeglichen. Von vorneherein ausgeglichene Datensätze wären hier jedoch der Datenaugmentierung vorzuziehen.

Dieser Arbeit liegt eine manuelle Segmentierung der ROIs zugrunde. Durch die möglichst genaue, manuelle Segmentierung und die anschließende Revision der ROIs durch erfahrene Radiologinnen und Radiologen war eine verlässliche „Ground Truth“ gegeben. In dieser Studie wurde die Prostata als gesamtes Organ, sowie die periphere Zone und Transitionalzone einzeln segmentiert. Dies wurde in den ADC-maps ebenfalls durchgeführt. Bei der Segmentierung wurde darauf geachtet, mit dem Linienverlauf möglichst exakt den anatomischen Randstrukturen zu entsprechen. Um dem System optimale Trainingsvoraussetzungen zu ermöglichen, wurde somit möglichst wenig biologisch andersartiges Gewebe in einem segmentierten Bereich eingeschlossen bzw. angeschnitten (z.B. möglichst wenig Kapselgewebe in den Randbereichen der TZ, oder möglichst wenig TZ-Gewebe in den Randbereichen der PZ). Das anteriore fibromuskuläre Stroma (AFMS) wurde bei der Segmentierung nicht ausgespart, sondern in die in die designierte „TZ“ miteingeschlossen.

Allerdings weist diese anatomische Region andere histologische und bildgebende Eigenschaften auf als die restliche Transitionalzone und die periphere Zone. Auch die Urethra wurde bei der Segmentierung nicht ausgespart, sondern mit in die „TZ“ eingeschlossen.

Daher könnten in diesem Fall heterogene Gewebe in den ROIs die Ergebnisgüte dieser Studie beeinträchtigt haben. Es wäre zu prüfen, ob eine strengere Aussparung der genannten Gewebestrukturen zu einem besseren Ergebnis führen würde.

Ein zonenspezifischer Ansatz wurde auch in dieser Studie exploriert. Der Ansatz wies in dieser Studie jedoch im Vergleich zum regulären schichtbasierten Ansatz nur eine eingeschränkte Aussagekraft auf. Eine Verwendung größerer Datensätze könnte hier ggf. stabilere Ergebnisse generieren.

Der Referenzstandard zur Validierung der KI-Algorithmen in dieser Studie ist die abschließende histologische Beurteilung von Stanzbiopsie- oder Prostatektomiepräparaten.

Jedoch gehen, wie in der Einleitung bereits dargestellt, die histopathologische Beurteilung von Biopsiepräparaten und die mpMRT mit einer relevanten Falsch-negativ-Rate insbesondere für kleine PCa einher. Die histologische Korrelation ist darüber hinaus auch insofern problematisch, als dass es ein übergeordnetes Ziel dieser Studie war, Patienten mit signifikantem PCa mittels nicht-invasiver, zuverlässiger bildgebender Diagnostik noch vor Biopsie oder Operation zu identifizieren.

Außerdem werden Prostatektomien als Therapiemethode mit zugleich diagnostischem Wert generell nur bei Vorliegen eines klinisch signifikanten und noch nicht metastasierten PCa durchgeführt, was bei der Verwendung als Referenzstandard ein Selektionsbias bedeutet.

Entsprechend differenziert muss der Referenzstandard dieser Studie betrachtet werden.

Darüber hinaus lag eine weitere Problematik dieser Studie darin, dass die Labels „kein PCa“ (benigne Läsion), „non-cs PCa“, und „cs PCa“, die durch Fachärztinnen und Fachärzte der Radiologie und Urologie im Konsens in schichtbasierter Vorgehensweise vergeben wurden, auf den histopathologischen Befunden basierten und visuell mit den mpMRT-Aufnahmen korreliert wurden. Jedoch besteht eines der Ziele eines Radiomics- bzw. Machine Learning-Algorithmus darin, eben gerade auch Läsionen zu detektieren, die dem menschlichen Auge verborgen bleiben. Die Annahme, dass kein (signifikantes) PCa vorliegt, weil in den mpMRT keines zu sehen ist, ist also problematisch, wie auch im Abschnitt „Undergrading“ der Einleitung bereits ausgeführt wurde.

Auch die herkömmliche Methode der PI-RADS-Klassifizierung eignet sich nur begrenzt als Korrelat, da diese idealerweise vom Algorithmus übertroffen werden soll. Außerdem bezieht sich das PI-RADS-Klassifikationssystem auf die gesamte Prostata, während diese Studie einen schichtbasierten Ansatz verfolgte. Dieser Problematik wurde jedoch die Verwendung des Sequence-Models entgegengesetzt, sodass ein Vergleich mit PI-RADS letztlich legitim und sinnvoll war.

Eine gesonderte externe Validierung fand in dieser Studie mangels externer ausgewerteter (d.h. mit Labels versehener) Daten nicht statt. Daher lässt sich über die Generalisierbarkeit der Modelle auf externe Daten keine Aussage treffen, was eine Limitation dieser Studie darstellt.

Mit DenseNet-121 wurde in dieser Studie ein erfolgreiches Deep Learning-Modell verwendet. Die Ergebnisse des hier alternativ zum Radiomics- bzw. Machine Learning-Ansatz angewandten Deep Learning-Ansatzes waren jedoch nicht zufriedenstellend (AUC von 0,59 +/- 0,04). Diesem Ergebnis steht zum Beispiel eine Studie entgegen, in der ein DL-Modell mit einem non-DL-Ansatz verglichen wurde und eine signifikant höhere Zuverlässigkeit und Genauigkeit (AUC 0,84 gegenüber 0,70) bei der Differenzierung zwischen PCa und benignen Veränderungen erreichte. Dieses Ergebnis konnte außerdem ohne vorherige manuelle Segmentierung eines Trainings-Sets erzielt werden [343]. Für das schlechtere Ergebnis dieser Studie können verschiedene Gründe diskutiert werden. Dieser Studie lag z.B. ein hochdimensionaler Datensatz bei relativ geringer Anzahl an Beispielen vor, zudem war die Aufgabenstellung in dieser Studie komplex. Das System hätte an einem größeren Datensatz

wahrscheinlich bessere Ergebnisse generieren können. Außerdem lag im Datensatz trotz der durchgeführten Datenaugmentierung eine Klassen-Imbalance vor, die auch die Analyse des hierfür empfindlichen Deep Learning-Systems negativ beeinflusst haben dürfte.

Der Einsatz des Sequence-Models beim Deep Learning-Ansatz verbesserte die Ergebnisse erheblich. Darüber hinaus ermöglichte er als „study-based approach“ den Vergleich mit PI-RADS und war deshalb ein großer Mehrwert für diese Studie. Jedoch muss einschränkend angemerkt werden, dass mit 33 Datensätzen für den Sequence-Model-Ansatz ein relativ kleiner Test-Datensatz vorlag. Darüber hinaus war auch in diesem Datensatz eine Klassen-Imbalance gegeben.

### 5.3.Fazit und Ausblick

Trotz fortgeschrittenem Verständnis für das PCa und moderner diagnostischer Möglichkeiten wie der mpMRT und zielgerichteter TRUS-Biopsie kommt es weiterhin zu Überdiagnose und konsekutiver Übertherapie, aber auch zu Undergrading von PCa-Erkrankungen.

Wie dargestellt, ist das PCa häufig indolent, sodass die möglichst genaue Beurteilung der Aggressivität von PCa von entscheidender Bedeutung für das Patientenmanagement ist [344]. Das Ziel ist es demnach, möglichst zuverlässig diejenigen Patienten mit Therapieindikation zu identifizieren und zugleich den hohen Anteil der indolenten Tumorerkrankungen nicht den Nachteilen einer Übertherapie auszusetzen.

Eine klinische Hauptproblematik in der Diagnostik des PCa besteht in der (möglichst auf nicht-invasiver Diagnostik basierenden) möglichst sensitiven und spezifischen Differenzierung zwischen klinisch signifikanten PCa und indolenten PCa, welche durch die Heterogenität und die diagnostische Komplexität der Erkrankung erschwert wird.

Diese Differenzierung konnte der in dieser Studie verfolgte Ansatz mit Erfolg realisieren. Wir konnten zeigen, dass ein Machine Learning- bzw. Deep Learning-Modell mit aus mpMRT extrahierten Features zwischen malignen und benignen Prostataläsionen differenzieren, und auch klinisch signifikante von nicht signifikanten PCa unterscheiden kann.

Der Ansatz, den Modellen die gesamte Prostata schichtbasiert zuzuführen, hat sich als vielversprechend erwiesen; die Klassifikationen auf Schicht-Ebene waren hochwertig.

Im Vergleich mit PI-RADS zeigt diese Studie, dass es durchaus möglich ist, mit einem vollautomatisierten Modell ohne vorherige, manuelle Läsionssegmentierungen in den Qualitätsbereich der menschlichen Befundung zu gelangen.

Die Anwendbarkeit dieser Methoden ist weiterhin Gegenstand der Forschung. Die klinische Übertragbarkeit von KI- bzw. DL-Methoden wurde bereits mehrfach untersucht [345-347].

Die Ergebnisse dieser Studie können dazu beitragen, einen zuverlässigen Machine Learning-Algorithmus zur Prädiktion klinisch signifikanter PCa für die klinische Anwendung zu entwickeln. Außerdem sollen die in dieser Studie gewonnenen Erfahrungen dazu beitragen, den Radiomics- bzw. Machine Learning-Ansatz in der klinischen Diagnostik besser zu verstehen und somit die Generierung zuverlässiger prognostischer Marker für das PCa voranzutreiben.

Die „Stand-alone“-Performance der KI-Algorithmen ist vielversprechend im Hinblick auf eine Anwendbarkeit im klinischen Setting. Jedoch sind Generalisierbarkeit, Robustheit und Reproduzierbarkeit zentrale Problemfelder in der aktuellen Studienlage und stehen der Übertragung in den klinischen Workflow im Wege [337].

Da die bisherige Studienlage zu einem Großteil aus retrospektiven, Single-Center und Single-Algorithmus-Studien besteht, die oft mit relativ kleinen Datensätzen durchgeführt wurden, ist die Beurteilbarkeit von Reproduzierbarkeit und Zuverlässigkeit der bisherigen Ergebnisse eingeschränkt. Insbesondere sollte künftig bei Studiendesign der Fokus auf größere

Kohortengrößen und externe Validierung gelegt werden. Es müssen perspektivisch große Datensätze mit guten Annotationen aus möglichst vielen verschiedenen Institutionen für prospektive Studien gewonnen werden, um Radiomics und KI weiter im Hinblick auf Nutzen und Implementierbarkeit für den klinischen Alltag zu prüfen.

Dabei wäre eine strukturierte Datenakquise und -archivierung als Grundlage wünschenswert. Laut Patel et al. ist die Etablierung strenger Leitlinien zu Studienplanung und Methodik nötig, um Qualität und Vergleichbarkeit der Forschungsergebnisse zu gewährleisten und zu optimieren [348].

Auch wenn die Studienergebnisse zum Teil divergieren, ist der Ausblick dennoch vielversprechend, da bei der zunehmenden Anwendung von Deep Learning eine stetige Leistungssteigerung der Systeme zu erwarten ist [249, 349].

Im Zuge der Digitalisierung werden weltweit immer mehr medizinische Bilddaten, aber auch digitalisierte klinische Informationen erzeugt, archiviert und damit zugänglich, welche das Potential einer „mineable resource“ („abbaubare Ressource“) innehaben. Dieses „data mining“ bietet die Möglichkeit zu aussagekräftigen retrospektiven Studien und könnte in der Zukunft einen exponentiellen Wissenszuwachs mit sich bringen [350]. Das The Cancer Imaging Archive (TCIA) ist hier erwähnenswert, das Informationen aus medizinischer Bildgebung, klinische und histologische Daten, bis hin zu genomischen Daten in einem Open-Access-Setting anbietet [351, 352].

Die reine Bildanalyse erzielt für sich genommen bereits gute Ergebnisse bei der Detektion und Charakterisierung von PCa. Dieses Potential lässt sich jedoch erweitern, führt man sich vor Augen, dass dieser Ansatz nicht in Gänze den klinischen Alltag abbildet: In der Realität werden die bildgebenden Informationen durch viele weitere hinweisende Informationen ergänzt, die von Ärztinnen und Ärzten zusätzlich zur Beurteilung eines Falls herangezogen werden. Hierzu zählen Ergebnisse weiterer Diagnostik wie DRU oder TRUS, PSA-Werte, PI-RADS- und Gleason Score, die Quantifizierung von Therapieansprechen, oder auch das Einbeziehen vergleichbarer Krankheitsverläufe oder Risikofaktoren wie Alter und familiäre Vorbelastung eines Patienten. Dementsprechend ist der schichtbasierte Ansatz dieser und vieler weiterer Studien zwar sehr gut objektivierbar und von invasiver Diagnostik unabhängig, er bildet aber unter diesem Gesichtspunkt nur ein Mosaik des klinischen Alltags ab.

Tatsächlich können ML- bzw. DL-Modelle verschiedenste Arten von Daten als Inputs verwenden und sind keineswegs auf die ausschließliche Verarbeitung von Radiomics Features limitiert. Die Radiomics-Analyse im Sinne eines „multiomics“-Modells durch klinische Parameter als Inputs zu ergänzen, um möglichst prädiktive Modelle zu entwickeln, stellt somit einen chancenreichen Ansatz dar, der im Hinblick auf PCa bereits vielversprechend untersucht wurde und wird [353].

Neben der Anwendung von Radiomics zur Detektion und Lokalisierung von PCa besteht die Möglichkeit, die Tumorbiologie (z.B. Aggressivität oder hypoxische Areale) auf genetischer Ebene einzuschätzen [95]. Auch können ML- und DL-gestützte CDS-Systeme laut mehrerer Studien bereits klinisch nicht signifikante PCa von signifikanten PCa differenzieren und GS 3 + 3 Läsionen von GS > 3 + 3 voneinander unterscheiden [269, 354, 355].

Die charakteristischen Auswirkungen einer Tumorerkrankung auf den Organismus stellen einen weiteren Ansatzpunkt für die Analyse durch Klassifikationsmodelle dar. Hierzu zählen beispielsweise Krankheitsmanifestationen wie bildmorphologisch auffällige Lymphknoten, ossäre Läsionen, Entzündungserscheinungen oder Nachbarorganinfiltration. Durch eine Vergrößerung des Analysefensters könnten derartige Informationen von einem Klassifikationsmodell zusätzlich erfasst und ausgewertet werden. Die reine prostatabezogene Analyse erzielt bereits vielversprechende Ergebnisse. Eine Ergänzung dieser Informationsgrundlage um solche relevante Daten, die mit derselben Untersuchung bereits erbracht wurden, birgt zusätzliches Potential für die automatisierte Bildanalyse im klinischen Workflow.

In den Jahren 2020 und 2021 haben mehrere Hersteller Zulassungen der Food and Drug Administration (FDA) für KI-Lösungen zur Prostata- bzw. PCa-Diagnostik erhalten [356-358]. Dies unterstreicht Anwendbarkeit und Potential der bildgebenden Prostatadiagnostik und der hier untersuchten Techniken.

Die künstliche Intelligenz hält nicht nur Einzug in die Radiologie, sondern in viele weitere Bereiche unseres Lebens. Manche Autoren sehen durch Radiomics und KI „eine neue Ära der Medizin“ eingeläutet [359-361]. Diese wird einige Probleme im gegenwärtigen klinischen Alltag lösen, aber auch neue Herausforderungen mit sich bringen.

Weiter Studien werden benötigt, um Radiomics, Machine Learning- und Deep Learning-Modelle in den klinischen Workflow von Diagnostik und individueller Risikostratifizierung von Patienten mit PCa zu implementieren.

## 6. ZUSAMMENFASSUNG

Das Prostatakarzinom ist weltweit die häufigste maligne Tumorerkrankung und die zweithäufigste tumorbezogene Todesursache des Mannes. Jedoch ist die Prävalenz indolenter Tumoren hoch, während das Sterberisiko in westlichen Industrieländern nur 2,5% beträgt.

Laut der Einschätzungen einer Reihe von Studien beläuft sich die Rate der Überdiagnose von klinisch nicht signifikanten Prostatakarzinomen auf bis zu 67%. Zudem steigen der Anteil falsch-positiver Ergebnisse im Screening und die Komplikationsrate bei Biopsien mit steigendem Lebensalter.

Die Therapieoptionen haben sich in den letzten Jahren zwar erweitert, können jedoch nach wie vor mit erheblich lebensqualitätseinschränkenden, irreversiblen Nebenwirkungen und Komplikationen einhergehen. Entsprechend sollte unbedingt vermieden werden, Patienten mit indolenten Karzinomen einer Therapie zu unterziehen („Übertherapie“).

Die erhebliche Diskrepanz zwischen hoher Inzidenz und Prävalenz einerseits und niedriger Mortalität andererseits begründet die dringende Notwendigkeit, bereits bei Diagnosestellung und auch im Verlauf sicher zwischen klinisch signifikanten Prostatakarzinomen mit Therapieindikation und indolenten Prostatakarzinomen zu differenzieren. Der individuellen Risikostratifizierung kommt auch vor dem Hintergrund eines Undergradings mit potentiell fatalen Folgen ein fundamentaler Stellenwert zu.

Die möglichst präzise Risikostratifizierung des Prostatakarzinoms ist außerdem Grundvoraussetzung zur Auswahl der bestmöglichen Therapiestrategie. Darüber hinaus hängt das Outcome eines Patienten entscheidend davon ab, wie akkurat prognoserelevante Faktoren wie Tumorlokalisierung, -ausdehnung, -volumen und -aggressivität prä-interventionell eingeschätzt werden.

Bisherige Prognosefaktoren wie Gleason-Score, PSA-Wert und TNM-Stadium gewährleisten nicht in ausreichendem Maße die präzise Charakterisierung von Prostatakarzinomen. Auch die digital-rektale Untersuchung und der transurethrale Ultraschall sind zur Beurteilung lokal fortgeschrittener Karzinome nicht akkurat genug. Z.B. wird bei bis zu 30% der Patienten mit extrakapsulärer Tumorausbreitung diese als bedeutsamer Prognosefaktor mit diesen Methoden nicht erfasst. Auch Lymphknotenmetastasen, die zu den für die Prognose relevantesten Faktoren gehören, werden nicht akkurat genug diagnostiziert. Selbst der Lymphadenektomie als Goldstandard zur Stagingdiagnostik entgehen bis zu 41% der Metastasen.

Der multiparametrischen MRT kommt in der Prostatakarzinom-Diagnostik eine zentrale Bedeutung zu. Durch die verschiedenen Bildgebungstechniken (T<sub>2</sub>w, DCE-MR, DWI) haben sich die Anwendungsmöglichkeiten der MRT erweitert. Sie liefert ohne die Nachteile von Strahlenbelastung oder Invasivität sowohl anatomisch-morphologisch als auch funktionell einen hohen Informationsgehalt. Jedoch werden durch die multiparametrische MRT große Datenmengen generiert, deren Interpretation im Sinne des im klinischen Alltag etablierten PI-RADS-Klassifikationssystems sehr komplex und zeitaufwändig sein kann. Zudem stellt neben einer geringen Sensitivität in Bezug auf eine Kapselinfiltration und einer signifikanten Unterschätzung des Tumolvolumens auch die Interreader-Variabilität bei der multiparametrischen MRT ein Problem dar. Auch kommt das Prostatakarzinom in äußerst heterogenen, oftmals multizentrischen Ausprägungen vor, sodass z.B. Mikrosatelliten-Herde nicht erfasst werden.

Diese Problematik begründet die Notwendigkeit neuer diagnostischer Methoden, die auf der einen Seite zwar sensitiv sind, auf der anderen Seite aber eine Risikostratifizierung priorisieren. Radiomics sind eine Methode der Bildanalyse, die quantitative Merkmale aus radiologischen Bilddatensätzen extrahiert und quantifiziert. Somit lassen sich bildphänotypische Merkmale ableiten, die als bildgebende Biomarker verstanden werden können. Zusammen mit der

Anwendung künstlicher Intelligenz (Machine Learning, Deep Learning) sollen Prädiktionen zur klinischen Signifikanz von Prostatakarzinomen möglich werden. Durch diese integrativen Analysealgorithmen soll die herkömmlicherweise sehr untersucherabhängige Befundung der multifaktoriellen Parameter zur genauen Charakterisierung von Prostatakarzinomen objektivierbar werden.

Ziel dieser Studie war es, durch die integrative Analyse bildgebender Parameter einen Analysealgorithmus zu entwickeln, der es ermöglicht, Patienten mit signifikantem Prostatakarzinom noch vor Biopsie oder Operation zu identifizieren.

Mit diesem Ziel wurden Methoden des Machine Learning bzw. Deep Learning in einem schichtbasierten Ansatz auf multiparametrische MRT-Aufnahmen angewendet. Dazu wurden die Schritte einer Radiomics- bzw. Machine Learning-Pipeline durchgeführt.

Im Rahmen einer retrospektiven Datenanalyse wurden 297 Patienten nach bestimmten Einschlusskriterien identifiziert.

Zunächst wurde in den T<sub>2</sub>w-Bildern und ADC-maps die Prostata manuell in die Zonen periphere Zone, Transitionalzone und das gesamte Organ segmentiert. Als Referenzstandard wurden die einzelnen Schichten der MRT durch Fachärztinnen und Fachärzte der Urologie und Radiologie konsensuell histopathologisch korreliert und Labels vergeben (benigne Läsion/indolentes PCa/signifikantes PCa). Die segmentierten Regions of Interest dienten als Input für die Extraktion von Radiomics Features der Klassen Shape-Features, First-Order-Statistics und Textur-Features.

Die Datensätze der Kohorte wurden in ein Trainings- und ein Validierungs-Datenset aufgeteilt. Es wurden verschiedene Machine-Learning-Modelle trainiert. Die Support Vector Machine erbrachte die beste Performance und wurde daher auf den Validierungs-Datensatz angewandt. Das Modell erreichte bei der Klassifikation des Trainings-Datensatzes in „benigne Läsion“ oder „Vorliegen eines PCa“ eine AUC von 0,88 und am Validierungs-Datensatz 0,86. Bei der Klassifikation in die beiden Kategorien „klinisch signifikantes PCa“ oder „benigne Läsion bzw. klinisch nicht signifikantes PCa“ wurde eine AUC von 0,79 (Trainings-Set) und 0,83 (Validierungs-Set) erzielt.

Außerdem wurde ein zonaler Ansatz exploriert, bei dem die periphere Zone und Transitionalzone getrennt voneinander analysiert wurden. Dieser Ansatz erbrachte in dieser Studie im Vergleich zum nicht-zonalen Ansatz schlechtere Ergebnisse (Klassifikation in „benigne Läsion“ oder „PCa“ nur in der peripheren Zone mit einer AUC von 0,75).

Alternativ zum Radiomics-Machine-Learning-Ansatz wurde ein Deep Learning-Ansatz exploriert. Bei schichtweiser Analyse war die Performance des Deep Learning-Klassifikationssystems nicht zufriedenstellend. Im zweiten Schritt wurde ein Sequence-Modell angewandt. Die Performance war mit einer AUC von 0,81 für die Klassifikation in „benigne Läsion“ oder „PCa“ besser als PI-RADS (AUC von 0,77), welches ein im klinischen Alltag fest etabliertes diagnostisches Klassifikationssystem ist. Auch bei der Differenzierung zwischen „klinisch nicht signifikantes PCa“ und „klinisch signifikantes PCa“ übertraf das Sequence-Modell die PI-RADS-Performance (AUC von 0,86 vs. 0,83).

Diese Studie konnte zeigen, dass aus multiparametrischen MRT prädiktive Radiomics Features abgeleitet werden können. Außerdem konnte gezeigt werden, dass es möglich ist, mit einem Radiomics-basierten Machine Learning-Modell oder auch einem automatisierten Deep Learning-Modell in den diagnostischen Qualitätsbereich des menschlichen Readers zu gelangen.

Die Ergebnisse dieser Studie können dazu beitragen, eine zuverlässige Radiomics-Signatur und einen Machine Learning- bzw. Deep Learning-Algorithmus zur Prädiktion klinisch signifikanter Prostatakarzinome für die klinische Anwendung zu entwickeln, der sich als Künstliche Intelligenz-Lösung zur verbesserten Risikostratifizierung von Patienten mit Prostatakarzinom in den klinischen Alltag implementieren ließe.

## 7. LITERATURVERZEICHNIS

1. Jemal, A., et al., *Cancer statistics, 2010*. CA Cancer J Clin, 2010. **60**(5): p. 277-300.
2. Pernar, C.H., et al., *The Epidemiology of Prostate Cancer*. Cold Spring Harb Perspect Med, 2018. **8**(12).
3. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394-424.
4. Bott, S.R., et al., *Prostate cancer management: (1) an update on localised disease*. Postgrad Med J, 2003. **79**(936): p. 575-80.
5. Grossman, D.C., et al., *Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement*. Jama, 2018. **319**(18): p. 1901-1913.
6. Dall'Era, M.A., et al., *Active surveillance for early-stage prostate cancer: review of the current literature*. Cancer, 2008. **112**(8): p. 1650-9.
7. Yatani, R., et al., *Geographic pathology of latent prostatic carcinoma*. Int J Cancer, 1982. **29**(6): p. 611-6.
8. Choi, J.B., et al., *Difference in prostate cancer incidence around sixty years: effects of age and metabolic diseases*. Cancer Med, 2018. **7**(6): p. 2736-2743.
9. Koch-Institut, R., *Krebs in Deutschland für 2015/2016 [Cancer in Germany in 2015/2016]*. 2019.
10. Winter, A., et al., *Zunahme des uroonkologischen Versorgungsbedarfs durch demographischen Wandel*. Der Urologe, 2015. **54**(9): p. 1261-1268.
11. Leitlinienprogramm Onkologie, *Leitlinienprogramm Onkologie: Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Langversion 5.1*. 2019, Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF.
12. Crawford, E.D., et al., *A retrospective analysis illustrating the substantial clinical and economic burden of prostate cancer*. Prostate Cancer Prostatic Dis, 2010. **13**(2): p. 162-7.
13. Zhou, C.K., et al., *Prostate cancer incidence in 43 populations worldwide: An analysis of time trends overall and by age group*. Int J Cancer, 2016. **138**(6): p. 1388-400.
14. Heijnsdijk, E.A., et al., *Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer*. Br J Cancer, 2009. **101**(11): p. 1833-8.
15. Gann, P.H., *Risk factors for prostate cancer*. Rev Urol, 2002. **4 Suppl 5**: p. S3-s10.
16. Ren, S., et al., *Whole-genome and Transcriptome Sequencing of Prostate Cancer Identify New Genetic Alterations Driving Disease Progression*. Eur Urol, 2018. **73**(3): p. 322-339.
17. Berger, M.F., et al., *The genomic complexity of primary human prostate cancer*. Nature, 2011. **470**(7333): p. 214-20.
18. Baca, S.C., et al., *Punctuated evolution of prostate cancer genomes*. Cell, 2013. **153**(3): p. 666-77.
19. Su, F., et al., *Spatial Intratumor Genomic Heterogeneity within Localized Prostate Cancer Revealed by Single-nucleus Sequencing*. Eur Urol, 2018. **74**(5): p. 551-559.
20. Damber, J.E., *Prostate cancer: epidemiology and risk factors*. Curr Opin Urol, 1998. **8**(5): p. 375-80.
21. Hemminki, K., J. Sundquist, and J.L. Bermejo, *How common is familial cancer?* Ann Oncol, 2008. **19**(1): p. 163-7.
22. Morote, J., I. Comas, and J. Planas, *Re: Nicolas Mottet, Joaquim Bellmunt, Erik Briers, et al. EAU-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer*. European Association

- of Urology; 2017. <http://uroweb.org/guideline/prostate-cancer>: *How to Assess the Efficacy of Medical Castration*. Eur Urol, 2018. **73**(5): p. e134-e135.
23. Westerman, M.E., et al., *Impact of a family history of prostate cancer on clinicopathologic outcomes and survival following radical prostatectomy*. World J Urol, 2016. **34**(8): p. 1115-22.
  24. Ramsey, S.D., et al., *On the importance of race, socioeconomic status and comorbidity when evaluating quality of life in men with prostate cancer*. J Urol, 2007. **177**(6): p. 1992-9.
  25. Pienta, K.J. and P.S. Esper, *Risk factors for prostate cancer*. Ann Intern Med, 1993. **118**(10): p. 793-803.
  26. Bratt, O., *Hereditary prostate cancer: clinical aspects*. J Urol, 2002. **168**(3): p. 906-13.
  27. Breslow, N., et al., *Latent carcinoma of prostate at autopsy in seven areas. The International Agency for Research on Cancer, Lyons, France*. Int J Cancer, 1977. **20**(5): p. 680-8.
  28. Koo, J.M. and B.S. Shim, *Significance of Serum Testosterone for Prostate-Specific Antigen (PSA) Elevation and Prediction of Prostate Cancer in Patients with PSA Above 10 ng/ml*. Korean J Urol, 2010. **51**(12): p. 831-5.
  29. Yassin, A., et al., *Testosterone, testosterone therapy and prostate cancer*. Aging Male, 2019. **22**(4): p. 219-227.
  30. Schünke M, S.U., Voll M, Wesker K, Schulte E. , *Prometheus Lernatlas - Innere Organe*. . 2 ed. 2009, Stuttgart - New York: Thieme.
  31. Boutros, P.C., et al., *Spatial genomic heterogeneity within localized, multifocal prostate cancer*. Nat Genet, 2015. **47**(7): p. 736-45.
  32. Sinnott, J.A., et al., *Molecular differences in transition zone and peripheral zone prostate tumors*. Carcinogenesis, 2015. **36**(6): p. 632-8.
  33. Davidson, D., et al., *Prostatic intraepithelial neoplasia is a risk factor for adenocarcinoma: predictive accuracy in needle biopsies*. J Urol, 1995. **154**(4): p. 1295-9.
  34. Humphrey, P.A., et al., *The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours*. Eur Urol, 2016. **70**(1): p. 106-119.
  35. Bostwick, D.G. and M.K. Brawer, *Prostatic intra-epithelial neoplasia and early invasion in prostate cancer*. Cancer, 1987. **59**(4): p. 788-94.
  36. Netto, G.J. and J.I. Epstein, *Widespread high-grade prostatic intraepithelial neoplasia on prostatic needle biopsy: a significant likelihood of subsequently diagnosed adenocarcinoma*. Am J Surg Pathol, 2006. **30**(9): p. 1184-8.
  37. Kramer, G., D. Mitteregger, and M. Marberger, *Is benign prostatic hyperplasia (BPH) an immune inflammatory disease?* Eur Urol, 2007. **51**(5): p. 1202-16.
  38. Packer, J.R. and N.J. Maitland, *The molecular and cellular origin of human prostate cancer*. Biochim Biophys Acta, 2016. **1863**(6 Pt A): p. 1238-60.
  39. Hudson, B.D., K.S. Kulp, and G.G. Loots, *Prostate cancer invasion and metastasis: insights from mining genomic data*. Brief Funct Genomics, 2013. **12**(5): p. 397-410.
  40. Lee, C.H., O. Akin-Olugbade, and A. Kirschenbaum, *Overview of prostate anatomy, histology, and pathology*. Endocrinol Metab Clin North Am, 2011. **40**(3): p. 565-75, viii-ix.
  41. Dago, A.E., et al., *Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells*. PLoS One, 2014. **9**(8): p. e101777.
  42. Humphrey, P.A., *Histopathology of Prostate Cancer*. Cold Spring Harb Perspect Med, 2017. **7**(10).

43. Gleason, D.F., *Classification of prostatic carcinomas*. Cancer Chemother Rep, 1966. **50**(3): p. 125-8.
44. Srigley, J.R., et al., *Updated protocol for the examination of specimens from patients with carcinomas of the prostate gland: a basis for checklists*. Cancer Committee. Arch Pathol Lab Med, 2000. **124**(7): p. 1034-9.
45. Albertsen, P.C., et al., *Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer*. Jama, 1998. **280**(11): p. 975-80.
46. Epstein, J.I., *An update of the Gleason grading system*. J Urol, 2010. **183**(2): p. 433-40.
47. Mahal, B.A., et al., *Gleason score 5 + 3 = 8 prostate cancer: much more like Gleason score 9?* BJU Int, 2016. **118**(1): p. 95-101.
48. Stark, J.R., et al., *Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3?* J Clin Oncol, 2009. **27**(21): p. 3459-64.
49. Schmelz HU, *Onkologie und Tumoren*, in *Facharztwissen Urologie*. 2006, Springer Medizin Verlag: Heidelberg.
50. Kristiansen, G., W. Roth, and B. Helpap, [*Grading of prostate cancer*]. Pathologe, 2016. **37**(4): p. 352-4.
51. Moch, H., P.A. Humphrey, and T.M. Ulbright, *Who classification of tumours of the urinary system and male genital organs*, in *WHO classification of tumours of the urinary system and male genital organs*. 2016. p. 356-356.
52. Pierorazio, P.M., et al., *Prognostic Gleason grade grouping: data based on the modified Gleason scoring system*. BJU international, 2013. **111**(5): p. 753-760.
53. Riede UN, Schaefer HE., *Allgemeine und spezielle Pathologie*. 1999, New York: Georg Thieme Verlag Stuttgart.
54. Wang, G., et al., *Genetics and biology of prostate cancer*. Genes Dev, 2018. **32**(17-18): p. 1105-1140.
55. Lin, S.C., L.Y. Yu-Lee, and S.H. Lin, *Osteoblastic Factors in Prostate Cancer Bone Metastasis*. Curr Osteoporos Rep, 2018. **16**(6): p. 642-647.
56. Herden, J., et al., *Predictive value of the UICC and AJCC 8th edition tumor-nodes-metastasis (TNM) classification for patients treated with radical prostatectomy*. Cancer Epidemiol, 2018. **56**: p. 126-132.
57. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer*. Jama, 1998. **280**(11): p. 969-74.
58. Thompson, I., et al., *Guideline for the management of clinically localized prostate cancer: 2007 update*. J Urol, 2007. **177**(6): p. 2106-31.
59. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018*. CA Cancer J Clin, 2018. **68**(1): p. 7-30.
60. Loeb, S., et al., *PSA doubling time versus PSA velocity to predict high-risk prostate cancer: data from the Baltimore Longitudinal Study of Aging*. Eur Urol, 2008. **54**(5): p. 1073-80.
61. Dunn, M.W., *Prostate Cancer Screening*. Semin Oncol Nurs, 2017. **33**(2): p. 156-164.
62. Hoedemaeker, R.F., et al., *Histopathological prostate cancer characteristics at radical prostatectomy after population based screening*. J Urol, 2000. **164**(2): p. 411-5.
63. Mistry, K. and G. Cable, *Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma*. J Am Board Fam Pract, 2003. **16**(2): p. 95-101.
64. Catalona, W.J., et al., *Comparison of Digital Rectal Examination and Serum Prostate Specific Antigen in the Early Detection of Prostate Cancer: Results of a Multicenter Clinical Trial of 6,630 Men*. J Urol, 2017. **197**(2s): p. S200-s207.

65. Stenman, U.H., et al., *Prostate-specific antigen*. *Semin Cancer Biol*, 1999. **9**(2): p. 83-93.
66. Greene, K.L., et al., *Prostate specific antigen best practice statement: 2009 update*. *J Urol*, 2013. **189**(1 Suppl): p. S2-s11.
67. Roobol, M.J., et al., *Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC)*. *Eur Urol*, 2009. **56**(4): p. 584-91.
68. Roobol, M.J., et al., *Screening for prostate cancer: results of the Rotterdam section of the European randomized study of screening for prostate cancer*. *Eur Urol*, 2013. **64**(4): p. 530-9.
69. Beyersdorff, D., et al., *Patients with a history of elevated prostate-specific antigen levels and negative transrectal US-guided quadrant or sextant biopsy results: value of MR imaging*. *Radiology*, 2002. **224**(3): p. 701-6.
70. Frauscher, F., et al., *[The value of ultrasound (US) in the diagnosis of prostate cancer]*. *Radiologe*, 2003. **43**(6): p. 455-63.
71. Shimizu, T., et al., *Prostate cancer detection: the value of performing an MRI before a biopsy*. *Acta Radiol*, 2009. **50**(9): p. 1080-8.
72. Halpern, E.J. and S.E. Strup, *Using gray-scale and color and power Doppler sonography to detect prostatic cancer*. *AJR Am J Roentgenol*, 2000. **174**(3): p. 623-7.
73. Heidenreich, A., et al., *EAU guidelines on prostate cancer*. *Eur Urol*, 2008. **53**(1): p. 68-80.
74. Carroll, P.R., et al., *NCCN Clinical Practice Guidelines Prostate Cancer Early Detection, Version 2.2015*. *J Natl Compr Canc Netw*, 2015. **13**(12): p. 1534-61.
75. Bootsma, A.M., et al., *Antibiotic prophylaxis in urologic procedures: a systematic review*. *Eur Urol*, 2008. **54**(6): p. 1270-86.
76. Westhoff, N., et al., *Systematic prostate biopsy still matters: A comprehensive analysis of MRI/TRUS-fusion targeted prostate biopsies across different indications*. *Urol Oncol*, 2019. **37**(10): p. 678-687.
77. Filson, C.P., et al., *Prostate cancer detection with magnetic resonance-ultrasound fusion biopsy: The role of systematic and targeted biopsies*. *Cancer*, 2016. **122**(6): p. 884-92.
78. Drost, F.H., et al., *Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer*. *Cochrane Database Syst Rev*, 2019. **4**: p. Cd012663.
79. Kuru, T.H., et al., *How to Perform Image-guided Prostate Biopsy: In-bore and Fusion Approaches*. *Eur Urol Focus*, 2016. **2**(2): p. 151-153.
80. Pepe, P., et al., *Is it Time to Perform Only Magnetic Resonance Imaging Targeted Cores? Our Experience with 1,032 Men Who Underwent Prostate Biopsy*. *J Urol*, 2018. **200**(4): p. 774-778.
81. Siddiqui, M.M., et al., *Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer*. *Jama*, 2015. **313**(4): p. 390-7.
82. Schoots, I.G., et al., *Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis*. *Eur Urol*, 2015. **68**(3): p. 438-50.
83. Rouviere, O., et al., *Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study*. *Lancet Oncol*, 2019. **20**(1): p. 100-109.

84. van der Leest, M., et al., *Head-to-head Comparison of Transrectal Ultrasound-guided Prostate Biopsy Versus Multiparametric Prostate Resonance Imaging with Subsequent Magnetic Resonance-guided Biopsy in Biopsy-naive Men with Elevated Prostate-specific Antigen: A Large Prospective Multicenter Clinical Study*. Eur Urol, 2019. **75**(4): p. 570-578.
85. Sertdemir, M., et al., *Is There a Role for Functional MRI for the Assessment of Extracapsular Extension in Prostate Cancer?* Anticancer Res, 2018. **38**(1): p. 427-432.
86. Hallscheidt, P. and A. Haferkamp, *Urogenitale bildgebung*. 2010: Springer-Verlag.
87. Novelline, R.A., *Squire's Radiologie: Grundlagen der klinischen Diagnostik für Studium und Praxis*. 2001: Schattauer Verlag.
88. Villers, A., et al., *Current status of MRI for the diagnosis, staging and prognosis of prostate cancer: implications for focal therapy and active surveillance*. Curr Opin Urol, 2009. **19**(3): p. 274-82.
89. Futterer, J.J., *Multiparametric MRI in the Detection of Clinically Significant Prostate Cancer*. Korean J Radiol, 2017. **18**(4): p. 597-606.
90. Rothke, M., et al., *[PI-RADS classification: structured reporting for MRI of the prostate]*. Rofo, 2013. **185**(3): p. 253-61.
91. Weinreb, J.C., et al., *PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2*. Eur Urol, 2016. **69**(1): p. 16-40.
92. Perera, M., et al., *An update on focal therapy for prostate cancer*. Nat Rev Urol, 2016. **13**(11): p. 641-653.
93. Schimmoller, L., et al., *MR-sequences for prostate cancer diagnostics: validation based on the PI-RADS scoring system and targeted MR-guided in-bore biopsy*. Eur Radiol, 2014. **24**(10): p. 2582-9.
94. Franiel, T., et al., *Areas suspicious for prostate cancer: MR-guided biopsy in patients with at least one transrectal US-guided biopsy with a negative finding--multiparametric MR imaging for detection and biopsy planning*. Radiology, 2011. **259**(1): p. 162-72.
95. Sun, Y., et al., *Multiparametric MRI and radiomics in prostate cancer: a review*. Australas Phys Eng Sci Med, 2019. **42**(1): p. 3-25.
96. Hoeks, C.M., et al., *Prostate cancer: multiparametric MR imaging for detection, localization, and staging*. Radiology, 2011. **261**(1): p. 46-66.
97. Sohayda, C., et al., *Extent of extracapsular extension in localized prostate cancer*. Urology, 2000. **55**(3): p. 382-6.
98. Engelbrecht, M.R., et al., *Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis*. Eur Radiol, 2002. **12**(9): p. 2294-302.
99. Turkbey, B., et al., *Imaging localized prostate cancer: current approaches and new developments*. AJR Am J Roentgenol, 2009. **192**(6): p. 1471-80.
100. Akin, O., et al., *Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging*. Radiology, 2006. **239**(3): p. 784-92.
101. Li, H., et al., *Conventional MRI capabilities in the diagnosis of prostate cancer in the transition zone*. AJR Am J Roentgenol, 2006. **186**(3): p. 729-42.
102. Le Bihan, D., *Molecular diffusion nuclear magnetic resonance imaging*. Magn Reson Q, 1991. **7**(1): p. 1-30.
103. Malayeri, A.A., et al., *Principles and applications of diffusion-weighted imaging in cancer detection, staging, and treatment follow-up*. Radiographics, 2011. **31**(6): p. 1773-91.
104. Meier-Schroers, M., et al., *Differentiation of prostatitis and prostate cancer using the Prostate Imaging-Reporting and Data System (PI-RADS)*. Eur J Radiol, 2016. **85**(7): p. 1304-11.

105. Oto, A., et al., *Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis*. AJR Am J Roentgenol, 2011. **197**(6): p. 1382-90.
106. Somford, D.M., et al., *Initial experience with identifying high-grade prostate cancer using diffusion-weighted MR imaging (DWI) in patients with a Gleason score  $\leq 3 + 3 = 6$  upon schematic TRUS-guided biopsy: a radical prostatectomy correlated series*. Invest Radiol, 2012. **47**(3): p. 153-8.
107. Bittencourt, L.K., et al., *Prostate MRI: diffusion-weighted imaging at 1.5T correlates better with prostatectomy Gleason Grades than TRUS-guided biopsies in peripheral zone tumours*. Eur Radiol, 2012. **22**(2): p. 468-75.
108. Barentsz, J.O., et al., *ESUR prostate MR guidelines 2012*. Eur Radiol, 2012. **22**(4): p. 746-57.
109. Wang, X., et al., *High-b-value diffusion-weighted MRI for the detection of prostate cancer at 3 T*. Clin Radiol, 2014. **69**(11): p. 1165-70.
110. Turkbey, B., et al., *Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2*. Eur Urol, 2019. **76**(3): p. 340-351.
111. Vargas, H.A., et al., *Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness*. Radiology, 2011. **259**(3): p. 775-84.
112. Kobus, T., et al., *Prostate cancer aggressiveness: in vivo assessment of MR spectroscopy and diffusion-weighted imaging at 3 T*. Radiology, 2012. **265**(2): p. 457-67.
113. Zelhof, B., et al., *Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer*. BJU Int, 2009. **103**(7): p. 883-8.
114. van As, N.J., et al., *A study of diffusion-weighted magnetic resonance imaging in men with untreated localised prostate cancer on active surveillance*. Eur Urol, 2009. **56**(6): p. 981-7.
115. Turkbey, B., et al., *Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images?* Radiology, 2011. **258**(2): p. 488-95.
116. Itou, Y., et al., *Clinical utility of apparent diffusion coefficient (ADC) values in patients with prostate cancer: can ADC values contribute to assess the aggressiveness of prostate cancer?* J Magn Reson Imaging, 2011. **33**(1): p. 167-72.
117. Pokharel, S.S., et al., *Multi-parametric MRI findings of transitional zone prostate cancers: correlation with 3-dimensional transperineal mapping biopsy*. Abdom Imaging, 2015. **40**(1): p. 143-50.
118. Park, S.Y., et al., *Diffusion-tensor MRI at 3 T: differentiation of central gland prostate cancer from benign prostatic hyperplasia*. AJR Am J Roentgenol, 2014. **202**(3): p. W254-62.
119. Shimofusa, R., et al., *Diffusion-weighted imaging of prostate cancer*. J Comput Assist Tomogr, 2005. **29**(2): p. 149-53.
120. Padhani, A.R., *Integrating multiparametric prostate MRI into clinical practice*. Cancer Imaging, 2011. **11 Spec No A**: p. S27-37.
121. Weinmann, H.J., et al., *Characteristics of gadolinium-DTPA complex: a potential NMR contrast agent*. AJR Am J Roentgenol, 1984. **142**(3): p. 619-24.
122. Hara, N., et al., *Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer*. Prostate, 2005. **62**(2): p. 140-7.

123. Ludemann, L., B. Hamm, and C. Zimmer, *Pharmacokinetic analysis of glioma compartments with dynamic Gd-DTPA-enhanced magnetic resonance imaging*. Magn Reson Imaging, 2000. **18**(10): p. 1201-14.
124. Sung, Y.S., et al., *Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps*. AJR Am J Roentgenol, 2011. **197**(5): p. 1122-9.
125. Engelbrecht, M.R., et al., *Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging*. Radiology, 2003. **229**(1): p. 248-54.
126. Weidner, N., et al., *Tumor angiogenesis correlates with metastasis in invasive prostate carcinoma*. Am J Pathol, 1993. **143**(2): p. 401-9.
127. Coakley, F.V. and H. Hricak, *Radiologic anatomy of the prostate gland: a clinical approach*. Radiol Clin North Am, 2000. **38**(1): p. 15-30.
128. Villers, A., A. Steg, and L. Boccon-Gibod, *Anatomy of the prostate: review of the different models*. Eur Urol, 1991. **20**(4): p. 261-8.
129. Puech, P., et al., *Prostate MRI: can we do without DCE sequences in 2013?* Diagn Interv Imaging, 2013. **94**(12): p. 1299-311.
130. Taghipour, M., et al., *Investigating the role of DCE-MRI, over T2 and DWI, in accurate PI-RADS v2 assessment of clinically significant peripheral zone prostate lesions as defined at radical prostatectomy*. Abdom Radiol (NY), 2019. **44**(4): p. 1520-1527.
131. Ullrich, T., et al., *Value of Dynamic Contrast-Enhanced (DCE) MR Imaging in Peripheral Lesions in PI-RADS-4 Patients*. Rofo, 2019.
132. Brancato, V., et al., *Assessment of DCE Utility for PCa Diagnosis Using PI-RADS v2.1: Effects on Diagnostic Accuracy and Reproducibility*. Diagnostics (Basel), 2020. **10**(3).
133. Hamm, C.A., et al., *[Artificial intelligence and radiomics in MRI-based prostate diagnostics]*. Radiologe, 2020. **60**(1): p. 48-55.
134. Röthke, M., et al. *PI-RADS-klassifikation: Strukturiertes befundungsschema für die MRT der prostata*. in *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. 2013. © Georg Thieme Verlag KG.
135. Thompson, J.E., et al., *Multiparametric magnetic resonance imaging guided diagnostic biopsy detects significant prostate cancer and could reduce unnecessary biopsies and over detection: a prospective study*. J Urol, 2014. **192**(1): p. 67-74.
136. Hötter, A. and O.F. Donati, *PI-RADS 2.1 und strukturierte Befundung der Magnetresonanztomographie der Prostata*. Der Radiologe, 2021. **61**(9): p. 802-809.
137. Malik, A., S. Srinivasan, and J. Batra, *A New Era of Prostate Cancer Precision Medicine*. Front Oncol, 2019. **9**: p. 1263.
138. Choo, R., et al., *PSA doubling time of prostate carcinoma managed with watchful observation alone*. Int J Radiat Oncol Biol Phys, 2001. **50**(3): p. 615-20.
139. Krakowsky, Y., A. Loblaw, and L. Klotz, *Prostate cancer death of men treated with initial active surveillance: clinical and biochemical characteristics*. J Urol, 2010. **184**(1): p. 131-5.
140. Parker, C., *Active surveillance: towards a new paradigm in the management of early prostate cancer*. Lancet Oncol, 2004. **5**(2): p. 101-6.
141. Roemeling, S., et al., *Active surveillance for prostate cancers detected in three subsequent rounds of a screening trial: characteristics, PSA doubling times, and outcome*. Eur Urol, 2007. **51**(5): p. 1244-50; discussion 1251.
142. König, I.R., et al., *What is precision medicine?* Eur Respir J, 2017. **50**(4).
143. Janiaud, P., S. Serghiou, and J.P.A. Ioannidis, *New clinical trial designs in the era of precision medicine: An overview of definitions, strengths, weaknesses, and current use in oncology*. Cancer Treat Rev, 2019. **73**: p. 20-30.

144. Calio, B.P., et al., *Spatial distribution of biopsy cores and the detection of intra-lesion pathologic heterogeneity*. *Ther Adv Urol*, 2019. **11**: p. 1756287219842485.
145. Mullane, S.A. and E.M. Van Allen, *Precision medicine for advanced prostate cancer*. *Curr Opin Urol*, 2016. **26**(3): p. 231-9.
146. Luboldt, H.J., et al., *[Early detection of prostatic carcinoma in urologic practice with digital rectal examination and prostate-specific antigen. Early Detection Project Group]*. *Urologe A*, 2000. **39**(4): p. 330-3.
147. Psutka, S.P., et al., *Men with organ-confined prostate cancer and positive surgical margins develop biochemical failure at a similar rate to men with extracapsular extension*. *Urology*, 2011. **78**(1): p. 121-5.
148. Schiavina, R., et al., *Identification of prostate cancer risk categories according to surgical margins status, pathological stage and Gleason score*. *Int J Urol*, 2013. **20**(11): p. 1097-103.
149. Shaw, G.L., et al., *Identification of pathologically insignificant prostate cancer is not accurate in unscreened men*. *Br J Cancer*, 2014. **110**(10): p. 2405-11.
150. Kojima, M., P. Troncoso, and R.J. Babaian, *Use of prostate-specific antigen and tumor volume in predicting needle biopsy grading error*. *Urology*, 1995. **45**(5): p. 807-12.
151. Thompson, I.M., et al., *Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter*. *N Engl J Med*, 2004. **350**(22): p. 2239-46.
152. Krumholtz, J.S., et al., *Prostate-specific antigen cutoff of 2.6 ng/mL for prostate cancer screening is associated with favorable pathologic tumor features*. *Urology*, 2002. **60**(3): p. 469-73; discussion 473-4.
153. Ohi, M., et al., *Diagnostic significance of PSA density adjusted by transition zone volume in males with PSA levels between 2 and 4ng/ml*. *Eur Urol*, 2004. **45**(1): p. 92-6; discussion 96-7.
154. Mullerad, M., et al., *Comparison of endorectal magnetic resonance imaging, guided prostate biopsy and digital rectal examination in the preoperative anatomical localization of prostate cancer*. *J Urol*, 2005. **174**(6): p. 2158-63.
155. de Rooij, M., et al., *Accuracy of Magnetic Resonance Imaging for Local Staging of Prostate Cancer: A Diagnostic Meta-analysis*. *Eur Urol*, 2016. **70**(2): p. 233-45.
156. Noguchi, M., et al., *Relationship between systematic biopsies and histological features of 222 radical prostatectomy specimens: lack of prediction of tumor significance for men with nonpalpable prostate cancer*. *J Urol*, 2001. **166**(1): p. 104-9; discussion 109-10.
157. D'Amico, A.V., et al., *Assessment of outcome prediction models for patients with localized prostate carcinoma managed with radical prostatectomy or external beam radiation therapy*. *Cancer*, 1998. **82**(10): p. 1887-96.
158. Epstein, J.I., et al., *Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer*. *Jama*, 1994. **271**(5): p. 368-74.
159. Bastian, P.J., et al., *Characteristics of insignificant clinical T1c prostate tumors. A contemporary analysis*. *Cancer*, 2004. **101**(9): p. 2001-5.
160. Futterer, J.J., et al., *Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature*. *Eur Urol*, 2015. **68**(6): p. 1045-53.
161. Djavan, B., et al., *Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3 and 4: when should we stop?* *J Urol*, 2001. **166**(5): p. 1679-83.
162. Smith, C.P. and B. Türkbey, *PI-RADS v2: Current standing and future outlook*. *Turk J Urol*, 2018. **44**(3): p. 189-194.
163. Bjurlin, M.A., J.S. Wysock, and S.S. Taneja, *Optimization of prostate biopsy: review of technique and complications*. *Urol Clin North Am*, 2014. **41**(2): p. 299-313.

164. Sheridan, T.B., et al., *Change in prostate cancer grade over time in men followed expectantly for stage T1c disease*. J Urol, 2008. **179**(3): p. 901-4; discussion 904-5.
165. Berglund, R.K., et al., *Pathological upgrading and up staging with immediate repeat biopsy in patients eligible for active surveillance*. J Urol, 2008. **180**(5): p. 1964-7; discussion 1967-8.
166. Alberts, A.R., et al., *Biopsy undergrading in men with Gleason score 6 and fatal prostate cancer in the European Randomized study of Screening for Prostate Cancer Rotterdam*. Int J Urol, 2017. **24**(4): p. 281-286.
167. Montironi, R., et al., *Handling of radical prostatectomy specimens: total embedding with large-format histology*. Int J Breast Cancer, 2012. **2012**: p. 932784.
168. Harbin, A.C. and D.D. Eun, *The role of extended pelvic lymphadenectomy with radical prostatectomy for high-risk prostate cancer*. Urol Oncol, 2015. **33**(5): p. 208-16.
169. Sathianathen, N.J., et al., *Accuracy of the magnetic resonance imaging pathway in the detection of prostate cancer: a systematic review and meta-analysis*. Prostate Cancer Prostatic Dis, 2019. **22**(1): p. 39-48.
170. Rosenkrantz, A.B., et al., *Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists*. Radiology, 2016. **280**(3): p. 793-804.
171. Rahbar, K., et al., *Correlation of Intraprostatic Tumor Extent with (6)(8)Ga-PSMA Distribution in Patients with Prostate Cancer*. J Nucl Med, 2016. **57**(4): p. 563-7.
172. NiMhurchu, E., et al., *Predictive value of PI-RADS classification in MRI-directed transrectal ultrasound guided prostate biopsy*. Clin Radiol, 2016. **71**(4): p. 375-80.
173. Greer, M.D., et al., *Accuracy and agreement of PIRADSV2 for prostate cancer mpMRI: A multireader study*. J Magn Reson Imaging, 2017. **45**(2): p. 579-585.
174. Rosenkrantz, A.B., et al., *Prostate Imaging Reporting and Data System (PI-RADS), Version 2: A Critical Look*. AJR Am J Roentgenol, 2016. **206**(6): p. 1179-83.
175. Mertan, F.V., et al., *Prospective Evaluation of the Prostate Imaging Reporting and Data System Version 2 for Prostate Cancer Detection*. J Urol, 2016. **196**(3): p. 690-6.
176. Muller, B.G., et al., *Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging*. Radiology, 2015. **277**(3): p. 741-50.
177. Park, S.Y., et al., *Prostate Cancer: PI-RADS Version 2 Helps Preoperatively Predict Clinically Significant Cancers*. Radiology, 2016. **280**(1): p. 108-16.
178. Rosenkrantz, A.B., et al., *Proposed Adjustments to PI-RADS Version 2 Decision Rules: Impact on Prostate Cancer Detection*. Radiology, 2017. **283**(1): p. 119-129.
179. Tan, N., et al., *In-Bore 3-T MR-guided Transrectal Targeted Prostate Biopsy: Prostate Imaging Reporting and Data System Version 2-based Diagnostic Performance for Detection of Prostate Cancer*. Radiology, 2017. **283**(1): p. 130-139.
180. Mehralivand, S., et al., *Prospective Evaluation of PI-RADS™ Version 2 Using the International Society of Urological Pathology Prostate Cancer Grade Group System*. J Urol, 2017. **198**(3): p. 583-590.
181. Stabile, A., et al., *Association Between Prostate Imaging Reporting and Data System (PI-RADS) Score for the Index Lesion and Multifocal, Clinically Significant Prostate Cancer*. Eur Urol Oncol, 2018. **1**(1): p. 29-36.
182. Jin, J., *Screening for Prostate Cancer*. Jama, 2018. **319**(18): p. 1946.
183. Zappa, M., et al., *Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study*. Ann Oncol, 1998. **9**(12): p. 1297-300.
184. Kilpeläinen, T.P., et al., *False-positive screening results in the European randomized study of screening for prostate cancer*. Eur J Cancer, 2011. **47**(18): p. 2698-705.

185. Fenton, J.J., et al., *Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force*. *Jama*, 2018. **319**(18): p. 1914-1931.
186. Porcaro, A.B., et al., *Low-Risk Prostate Cancer and Tumor Upgrading in the Surgical Specimen: Analysis of Clinical Factors Predicting Tumor Upgrading in a Contemporary Series of Patients Who were Evaluated According to the Modified Gleason Score Grading System*. *Curr Urol*, 2017. **10**(3): p. 118-125.
187. Mols, F., et al., *Bowel, urinary, and sexual problems among long-term prostate cancer survivors: a population-based study*. *Int J Radiat Oncol Biol Phys*, 2009. **73**(1): p. 30-8.
188. Stanford, J.L., et al., *Urinary and sexual function after radical prostatectomy for clinically localized prostate cancer: the Prostate Cancer Outcomes Study*. *Jama*, 2000. **283**(3): p. 354-60.
189. Etzioni, R., et al., *Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends*. *J Natl Cancer Inst*, 2002. **94**(13): p. 981-90.
190. Morash, C., et al., *Active surveillance for the management of localized prostate cancer: Guideline recommendations*. *Can Urol Assoc J*, 2015. **9**(5-6): p. 171-8.
191. Prensner, J.R., et al., *Beyond PSA: the next generation of prostate cancer biomarkers*. *Sci Transl Med*, 2012. **4**(127): p. 127rv3.
192. Collins, G.N., et al., *The effect of digital rectal examination, flexible cystoscopy and prostatic biopsy on free and total prostate specific antigen, and the free-to-total prostate specific antigen ratio in clinical practice*. *J Urol*, 1997. **157**(5): p. 1744-7.
193. Di Silverio, F., et al., *Relationship among age, prostate-specific antigen, and prostate volume in men with lower urinary tract symptoms (LUTS) and in different groups of men with and without benign and malignant prostate diseases*. *Prostate*, 1998. **36**(1): p. 1-7.
194. Tsivian, M., et al., *Predicting unilateral prostate cancer on routine diagnostic biopsy: sextant vs extended*. *BJU Int*, 2010. **105**(8): p. 1089-92.
195. Venderink, W., et al., *Results of Targeted Biopsy in Men with Magnetic Resonance Imaging Lesions Classified Equivocal, Likely or Highly Likely to Be Clinically Significant Prostate Cancer*. *Eur Urol*, 2018. **73**(3): p. 353-360.
196. Pathirana, T., et al., *Lifetime risk of prostate cancer overdiagnosis in Australia: quantifying the risk of overdiagnosis associated with prostate cancer screening in Australia using a novel lifetime risk approach*. *BMJ Open*, 2019. **9**(3): p. e022457.
197. Heijnsdijk, E.A., et al., *Quality-of-life effects of prostate-specific antigen screening*. *N Engl J Med*, 2012. **367**(7): p. 595-605.
198. Welch, H.G. and W.C. Black, *Overdiagnosis in cancer*. *J Natl Cancer Inst*, 2010. **102**(9): p. 605-13.
199. Loeb, S., et al., *Overdiagnosis and overtreatment of prostate cancer*. *Eur Urol*, 2014. **65**(6): p. 1046-55.
200. Radtke, R., *Radiologen - Anzahl in Deutschland nach ärztlichen Tätigkeitsbereichen 2019*. 2020: statista.
201. Demircioğlu, A., *[Radiomics-AI-based image analysis]*. *Pathologe*, 2019. **40**(Suppl 3): p. 271-276.
202. Bi, W.L., et al., *Artificial intelligence in cancer imaging: Clinical challenges and applications*. *CA Cancer J Clin*, 2019. **69**(2): p. 127-157.
203. Cuocolo, R., et al., *Machine learning applications in prostate cancer magnetic resonance imaging*. *Eur Radiol Exp*, 2019. **3**(1): p. 35.
204. Aerts, H.J., et al., *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach*. *Nat Commun*, 2014. **5**: p. 4006.

205. Grapov, D., et al., *Rise of Deep Learning for Genomic, Proteomic, and Metabolomic Data Integration in Precision Medicine*. Omics, 2018. **22**(10): p. 630-636.
206. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data*. Radiology, 2016. **278**(2): p. 563-77.
207. Murray, J.M., et al., [A primer on radiomics]. Radiologe, 2020. **60**(1): p. 32-41.
208. Segal, E., et al., *Decoding global gene expression programs in liver cancer by noninvasive imaging*. Nat Biotechnol, 2007. **25**(6): p. 675-80.
209. Lambin, P., et al., *Radiomics: extracting more information from medical images using advanced feature analysis*. Eur J Cancer, 2012. **48**(4): p. 441-6.
210. Rios Velazquez, E., et al., *Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer*. Cancer Res, 2017. **77**(14): p. 3922-3930.
211. Bodalal, Z., et al., *Radiogenomics: bridging imaging and genomics*. Abdom Radiol (NY), 2019. **44**(6): p. 1960-1984.
212. Patyk, M., et al., *Radiomics - the value of the numbers in present and future radiology*. Pol J Radiol, 2018. **83**: p. e171-e174.
213. Parekh, V.S. and M.A. Jacobs, *Deep learning and radiomics in precision medicine*. Expert Rev Precis Med Drug Dev, 2019. **4**(2): p. 59-72.
214. Kreit, E., et al., *Biological versus electronic adaptive coloration: how can one inform the other?* J R Soc Interface, 2013. **10**(78): p. 20120601.
215. O'Connor, J.P., et al., *Imaging biomarker roadmap for cancer studies*. Nat Rev Clin Oncol, 2017. **14**(3): p. 169-186.
216. Abbasi, A.N., A.M.H. Khan, and B.M. Qureshi, *Radiomics - Using Artificial Intelligence in the Quest towards Personalised Radiation Treatment*. J Coll Physicians Surg Pak, 2019. **29**(2): p. 99-100.
217. Rizzo, S., et al., *CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer*. Eur Radiol, 2016. **26**(1): p. 32-42.
218. Larue, R., et al., *Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study*. Acta Oncol, 2017. **56**(11): p. 1544-1553.
219. Rizzo, S., et al., *Radiomics: the facts and the challenges of image analysis*. Eur Radiol Exp, 2018. **2**(1): p. 36.
220. Haralick RM, S.K., Dinstein IH, *Textural features for image classification*. IEEE Trans Syst Man Cybern, 1973.
221. Mari, G.T.B.F.C.N.S.P.P.C.N.L.J.S.J.-L., *Texture Indexes and Gray Level Size Zone Matrix. Application to Cell Nuclei Classification*. Pattern Recognition and Information Processing (PRIP), 2009. **140-145**.
222. Hamid R Arabnia, Q.N.T., *Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology*. 1 ed. Systems and Applications, ed. H.R.A. Quoc Nam Tran. 2016: Morgan Kaufmann.
223. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. 2005. Ieee.
224. Ojala, T., M. Pietikainen, and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on pattern analysis and machine intelligence, 2002. **24**(7): p. 971-987.
225. Collewet, G., M. Strzelecki, and F. Mariette, *Influence of MRI acquisition protocols and image intensity normalization methods on texture classification*. Magnetic Resonance Imaging, 2004. **22**(1): p. 81-91.
226. Emmert-Streib, F., O. Yli-Harja, and M. Dehmer, *Artificial Intelligence: A clarification of misconceptions, myths and desired status*. Frontiers in Artificial Intelligence, 2020. **3**.

227. Döbel, L., Vogelsang *Maschinelles Lernen*, F.-G.z.F.r.d.a.F. e.V., Editor. 2018: München. p. [https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer\\_Studie\\_ML\\_201809.pdf](https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf).
228. Herold, C.J. and S. Delorme, [*Brave new world*]. *Radiologe*, 2020. **60**(1): p. 1-5.
229. Guo, Y., Y. Gao, and D. Shen, *Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching*. *IEEE Trans Med Imaging*, 2016. **35**(4): p. 1077-89.
230. I. H. Witten, E.F., and M. A. Hall *Data Mining: Practical Machine Learning Tools and Techniques*. 2011.
231. Breiman, L., et al., *Classification and regression trees*. 1984: CRC press.
232. Hackeling, G., *Mastering Machine Learning with scikit-learn*. 2014: Packt Publishing.
233. Breiman, *Random Forests*. *Machine Learning*, 2001. **45.1**.
234. Wang, Q.Q., et al., [*Overview of logistic regression model analysis and application*]. *Zhonghua Yu Fang Yi Xue Za Zhi*, 2019. **53**(9): p. 955-960.
235. Stoltzfus, J.C., *Logistic regression: a brief primer*. *Acad Emerg Med*, 2011. **18**(10): p. 1099-104.
236. Huang, S., et al., *Applications of Support Vector Machine (SVM) Learning in Cancer Genomics*. *Cancer Genomics Proteomics*, 2018. **15**(1): p. 41-51.
237. Afifi, S., H. GholamHosseini, and R. Sinha, *SVM classifier on chip for melanoma detection*. *Conf Proc IEEE Eng Med Biol Soc*, 2017. **2017**: p. 270-274.
238. Wu, M., et al., *Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis*. *J Magn Reson Imaging*, 2019. **50**(3): p. 940-950.
239. Varghese, B., et al., *Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images*. *Scientific reports*, 2019. **9**(1).
240. Tianqi Chen, Carlos Guestrin, *XGBoost: A Scalable Tree Boosting System*. 2016.
241. The XGBoost Contributors, *dmlc XGBoost*. 2014.
242. Schapire, R., *The strength of weak learnability*. *Machine Learning*. Vol. 5. 1990.
243. Hastie, , Tibshirani,, Friedman, *Data Mining, Inference, and Prediction, Second Edition*. 2 ed. *The Elements of Statistical Learning*. 2009, New York: Springer Verlag.
244. Zhang, Z., et al., *Predictive analytics with gradient boosting in clinical medicine*. *Ann Transl Med*, 2019. **7**(7): p. 152.
245. Livne, M., et al., *Boosted Tree Model Reforms Multimodal Magnetic Resonance Imaging Infarct Prediction in Acute Stroke*. *Stroke*, 2018. **49**(4): p. 912-918.
246. Ji, X., et al., *Five-Feature Model for Developing the Classifier for Synergistic vs. Antagonistic Drug Combinations Built by XGBoost*. *Front Genet*, 2019. **10**: p. 600.
247. Li, W., et al., *Gene Expression Value Prediction Based on XGBoost Algorithm*. *Front Genet*, 2019. **10**: p. 1077.
248. Kooi, T., et al., *Large scale deep learning for computer aided detection of mammographic lesions*. *Med Image Anal*, 2017. **35**: p. 303-312.
249. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks*. *Nature*, 2017. **542**(7639): p. 115-118.
250. Haenssle, H.A., et al., *Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists*. *Ann Oncol*, 2018. **29**(8): p. 1836-1842.
251. Jamaludin, A., T. Kadir, and A. Zisserman, *SpineNet: Automated classification and evidence visualization in spinal MRIs*. *Med Image Anal*, 2017. **41**: p. 63-73.

252. Gulshan, V., et al., *Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs*. *Jama*, 2016. **316**(22): p. 2402-2410.
253. Poplin, R., et al., *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning*. *Nat Biomed Eng*, 2018. **2**(3): p. 158-164.
254. Langs, G., et al., *Machine learning: from radiomics to discovery and routine*. *Radiologe*, 2018. **58**(Suppl 1): p. 1-6.
255. Kleesiek, J., et al., *[A primer on machine learning]*. *Radiologe*, 2020. **60**(1): p. 24-31.
256. McBee, M.P., et al., *Deep Learning in Radiology*. *Acad Radiol*, 2018. **25**(11): p. 1472-1480.
257. Greer, M.D., et al., *Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study*. *Eur Radiol*, 2018. **28**(10): p. 4407-4417.
258. Aerts, H., *Data Science in Radiology: A Path Forward*. *Clin Cancer Res*, 2018. **24**(3): p. 532-534.
259. Sullivan, D.C., et al., *Metrology Standards for Quantitative Imaging Biomarkers*. *Radiology*, 2015. **277**(3): p. 813-25.
260. Zwanenburg, A., et al., *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping*. *Radiology*, 2020: p. 191145.
261. Mackin, D., et al., *Measuring Computed Tomography Scanner Variability of Radiomics Features*. *Invest Radiol*, 2015. **50**(11): p. 757-65.
262. Baessler, B., K. Weiss, and D. Pinto Dos Santos, *Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study*. *Invest Radiol*, 2019. **54**(4): p. 221-228.
263. Steenbergen, P., et al., *Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation*. *Radiother Oncol*, 2015. **115**(2): p. 186-90.
264. Kleesiek, J., et al., *Virtual Raters for Reproducible and Objective Assessments in Radiology*. *Sci Rep*, 2016. **6**: p. 25007.
265. Isensee, F., et al., *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation*. *Nature Methods*, 2021. **18**(2): p. 203-211.
266. Kostyszyn, D., et al., *Intraprostatic Tumor Segmentation on PSMA PET Images in Patients with Primary Prostate Cancer with a Convolutional Neural Network*. *J Nucl Med*, 2021. **62**(6): p. 823-828.
267. Wang, B., et al., *Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation*. *Med Phys*, 2019. **46**(4): p. 1707-1718.
268. Yan, K., et al., *A propagation-DNN: Deep combination learning of multi-level features for MR prostate segmentation*. *Comput Methods Programs Biomed*, 2019. **170**: p. 11-21.
269. Antonelli, M., et al., *Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists*. *Eur Radiol*, 2019. **29**(9): p. 4754-4764.
270. Ginsburg, S.B., et al., *Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study*. *J Magn Reson Imaging*, 2017. **46**(1): p. 184-193.
271. Moradmand, H., S.M.R. Aghamiri, and R. Ghaderi, *Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma*. *J Appl Clin Med Phys*, 2020. **21**(1): p. 179-190.

272. Malley, B., D. Ramazzotti, and J.T.-y. Wu, *Data Pre-processing*, in *Secondary Analysis of Electronic Health Records*. 2016, Springer International Publishing: Cham. p. 115-141.
273. Isaksson, L.J., et al., *Effects of MRI image normalization techniques in prostate cancer radiomics*. *Physica Medica*, 2020. **71**: p. 7-13.
274. Scalco, E., et al., *T2w-MRI signal normalization affects radiomics features reproducibility*. *Medical physics*, 2020. **47**(4): p. 1680-1691.
275. Keenan, K.E., et al., *Challenges in ensuring the generalizability of image quantitation methods for MRI*. *Med Phys*, 2021.
276. Collewet, G., M. Strzelecki, and F. Mariette, *Influence of MRI acquisition protocols and image intensity normalization methods on texture classification*. *Magn Reson Imaging*, 2004. **22**(1): p. 81-91.
277. van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L., *Computational Radiomics System to Decode the Radiographic Phenotype*. *Cancer Research*. 2017.
278. Zwanenburg, A., et al., *The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping*. *Radiology*, 2020. **295**(2): p. 328-338.
279. Jha, A.K., et al., *Repeatability and reproducibility study of radiomic features on a phantom and human cohort*. *Sci Rep*, 2021. **11**(1): p. 2055.
280. Shiri, I., et al., *Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: Test-retest and image registration analyses*. *Med Phys*, 2020. **47**(9): p. 4265-4280.
281. Chawla, N., et al., *SMOTE: Synthetic Minority Over-sampling Technique*. *J. Artif. Intell. Res. (JAIR)*, 2002. **16**: p. 321-357.
282. Shafiq-Ul-Hassan, M., et al., *Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels*. *Med Phys*, 2017. **44**(3): p. 1050-1062.
283. Götz, M., M. Nolden, and K. Maier-Hein, *MITK Phenotyping: An open-source toolchain for image-based personalized medicine with radiomics*. *Radiother Oncol*, 2019. **131**: p. 108-111.
284. van Griethuysen, J.J.M., et al., *Computational Radiomics System to Decode the Radiographic Phenotype*. *Cancer Res*, 2017. **77**(21): p. e104-e107.
285. Nioche, C., et al., *LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity*. *Cancer Res*, 2018. **78**(16): p. 4786-4789.
286. Michael Götz, M.N., Klaus Maier-Hein, *MITK Phenotyping: An open-source toolchain for image-based personalized medicine with radiomics*. 2018, Radiotherapy and Oncology.
287. Parmar, C., et al., *Machine Learning methods for Quantitative Radiomic Biomarkers*. *Sci Rep*, 2015. **5**: p. 13087.
288. Leger, S., et al., *A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling*. *Sci Rep*, 2017. **7**(1): p. 13206.
289. Orlhac, F., et al., *Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis*. *J Nucl Med*, 2014. **55**(3): p. 414-22.
290. Guezennec, C., et al., *Inter-observer and segmentation method variability of textural analysis in pre-therapeutic FDG PET/CT in head and neck cancer*. *PLoS One*, 2019. **14**(3): p. e0214299.
291. Parmar, C., et al., *Robust Radiomics feature quantification using semiautomatic volumetric segmentation*. *PLoS One*, 2014. **9**(7): p. e102107.

292. Fournier, L., et al. *Incorporating radiomics into clinical trials: expert consensus on considerations for data-driven compared to biologically driven quantitative biomarkers*. in *European Radiology: journal of the European Congress of Radiology*. 2021.
293. Kohavi, R., *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Vol. 14. 1995.
294. Fawcett, *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*. *Pattern Recognition Letters*, 2004. **8**: p. S. 1–38.
295. Baeßler, B., K. Weiss, and D.P. Dos Santos, *Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study*. *Investigative radiology*, 2019. **54**(4): p. 221-228.
296. Mayerhoefer, M.E., et al., *Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study*. *Investigative radiology*, 2009. **44**(7): p. 405-411.
297. Eck, B., et al., *Prospective Evaluation of Repeatability and Robustness of Radiomic Descriptors in Healthy Brain Tissue Regions In Vivo Across Systematic Variations in T2-Weighted Magnetic Resonance Imaging Acquisition Parameters*. *Journal of Magnetic Resonance Imaging*, 2021.
298. Foster, K.R., R. Koprowski, and J.D. Skufca, *Machine learning, medical diagnosis, and biomedical engineering research-commentary*. *Biomedical engineering online*, 2014. **13**(1): p. 1-9.
299. Fiset, S., et al., *Repeatability and reproducibility of MRI-based radiomic features in cervical cancer*. *Radiotherapy and Oncology*, 2019. **135**: p. 107-114.
300. Mi, H., et al., *Impact of different scanners and acquisition parameters on robustness of MR radiomics features based on women's cervix*. *Scientific reports*, 2020. **10**(1): p. 1-12.
301. Sanduleanu, S., et al., *Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score*. *Radiotherapy and Oncology*, 2018. **127**(3): p. 349-360.
302. Mongan, J., L. Moy, and C.E. Kahn Jr, *Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers*. 2020, Radiological Society of North America.
303. Park, J.E., et al., *Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement*. *European radiology*, 2020. **30**(1): p. 523-536.
304. Stanzione, A., et al., *Prostate MRI radiomics: A systematic review and radiomic quality score assessment*. *Eur J Radiol*, 2020. **129**: p. 109095.
305. Litjens, G., et al., *A survey on deep learning in medical image analysis*. *Med Image Anal*, 2017. **42**: p. 60-88.
306. Zaharchuk, G., et al., *Deep Learning in Neuroradiology*. *AJNR Am J Neuroradiol*, 2018. **39**(10): p. 1776-1784.
307. Esteva, A., et al., *A guide to deep learning in healthcare*. *Nat Med*, 2019. **25**(1): p. 24-29.
308. Y. LeCun et al., *Deep Learning*. *Nature*, 2015. **521**: p. pp. 436– 444.
309. Srinivas et al., *Chapter 2 - An Introduction to Deep Convolutional Neural Nets for Computer Vision*. *Deep Learning for Medical Image Analysis*, 2017: p. pp. 26–30.
310. Diagnostic Image Analysis Group, *Grand Challenges in Biomedical Image Analysis*. 2019. p. <https://grand-challenge.org>.
311. Moeskops, P., et al., *Automatic Segmentation of MR Brain Images With a Convolutional Neural Network*. *IEEE Trans Med Imaging*, 2016. **35**(5): p. 1252-1261.

312. Vaillant, R., C. Monrocq, and Y. Le Cun, *Original approach for the localisation of objects in images*. IEE Proceedings-Vision, Image and Signal Processing, 1994. **141**(4): p. 245-250.
313. Ciregan, D., U. Meier, and J. Schmidhuber. *Multi-column deep neural networks for image classification*. in *2012 IEEE conference on computer vision and pattern recognition*. 2012. IEEE.
314. Lowekamp, B.C., et al., *The design of SimpleITK*. Frontiers in neuroinformatics, 2013. **7**: p. 45.
315. Yaniv, Z., et al., *SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research*. Journal of digital imaging, 2018. **31**(3): p. 290-303.
316. Peng, H., F. Long, and C. Ding, *Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy*. IEEE Transactions on pattern analysis and machine intelligence, 2005. **27**(8): p. 1226-1238.
317. Zwanenburg A, Leger S., Vallières M, Löck S *Image bio- marker standardisation initiative*. Radiology (2020), 2016.
318. Cournapeau, D., *scikit-learn*, in *Google Summer of Code*. 2010.
319. Allaire, *RStudio*. 2009: Boston, USA.
320. MONAI, *Project MONAI*. 2020: London.
321. Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
322. Liu, S., et al., *Prostate cancer diagnosis using deep learning with 3D multiparametric MRI*. SPIE Medical Imaging. Vol. 10134. 2017: SPIE.
323. Pan, I., *Deep Learning for Pulmonary Embolism Detection: Tackling the RSNA 2020 AI Challenge*. Radiology: Artificial Intelligence, 2021: p. e210068.
324. Zhang, L., et al., *Development and validation of a multiparametric MRI-based radiomics signature for distinguishing between indolent and aggressive prostate cancer*. Br J Radiol, 2021: p. 20210191.
325. Hectors, S.J., et al., *Magnetic Resonance Imaging Radiomics-Based Machine Learning Prediction of Clinically Significant Prostate Cancer in Equivocal PI-RADS 3 Lesions*. J Magn Reson Imaging, 2021.
326. Hou, Y., et al., *A radiomics machine learning-based redefining score robustly identifies clinically significant prostate cancer in equivocal PI-RADS score 3 lesions*. Abdom Radiol (NY), 2020. **45**(12): p. 4223-4234.
327. Zhang, Y., et al., *Development of a Novel, Multi-Parametric, MRI-Based Radiomic Nomogram for Differentiating Between Clinically Significant and Insignificant Prostate Cancer*. Front Oncol, 2020. **10**: p. 888.
328. Li, M., et al., *Radiomics prediction model for the improved diagnosis of clinically significant prostate cancer on biparametric MRI*. Quant Imaging Med Surg, 2020. **10**(2): p. 368-379.
329. Castaldo, R., et al., *Radiomic and Genomic Machine Learning Method Performance for Prostate Cancer Diagnosis: Systematic Literature Review*. J Med Internet Res, 2021. **23**(4): p. e22394.
330. Wang, J., et al., *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**(10): p. 4082-4090.
331. Nketiah, G.A., et al., *Utility of T(2)-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study*. Sci Rep, 2021. **11**(1): p. 2085.

332. Bleker, J., et al., *Multiparametric MRI and auto-fixed volume of interest-based radiomics signature for clinically significant peripheral zone prostate cancer*. Eur Radiol, 2020. **30**(3): p. 1313-1324.
333. Sidhu, H.S., et al., *"Textural analysis of multiparametric MRI detects transition zone prostate cancer"*. Eur Radiol, 2017. **27**(6): p. 2348-2358.
334. Chen, Q., et al. *A transfer learning approach for classification of clinical significant prostate cancers from mpMRI scans*. in *Medical Imaging 2017: Computer-Aided Diagnosis*. 2017. International Society for Optics and Photonics.
335. Ginsburg, S.B., et al., *Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study*. J Magn Reson Imaging, 2017. **46**(1): p. 184-193.
336. Schelb, P., et al., *Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment*. Radiology, 2019. **293**(3): p. 607-617.
337. Twilt, J.J., et al., *Artificial Intelligence Based Algorithms for Prostate Cancer Classification and Detection on Magnetic Resonance Imaging: A Narrative Review*. Diagnostics (Basel), 2021. **11**(6).
338. Brunese, L., et al., *Radiomics for Gleason Score Detection through Deep Learning*. Sensors (Basel), 2020. **20**(18).
339. Le, M.H., et al., *Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks*. Physics in Medicine & Biology, 2017. **62**(16): p. 6497.
340. Miotto, R., et al., *Deep learning for healthcare: review, opportunities and challenges*. Brief Bioinform, 2018. **19**(6): p. 1236-1246.
341. Thomas G. Tape, *Interpreting Diagnostic Tests*. University of Nebraska Medical Center. p. <http://gim.unmc.edu/dxtests/roc3.htm>.
342. American College of Radiology. *PI-RADS v2.1*. . 2019 [cited 2020 06.04.2020]; Available from: <https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2-1.pdf>.
343. Wang, X., et al., *Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning*. Sci Rep, 2017. **7**(1): p. 15415.
344. Mottet, N., et al., *EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent*. Eur Urol, 2017. **71**(4): p. 618-629.
345. Deniffel, D., et al., *Using decision curve analysis to benchmark performance of a magnetic resonance imaging-based deep learning model for prostate cancer risk assessment*. Eur Radiol, 2020. **30**(12): p. 6867-6876.
346. Song, Y., et al., *Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI*. J Magn Reson Imaging, 2018. **48**(6): p. 1570-1577.
347. Takeuchi, T., et al., *Prediction of prostate cancer by deep learning with multilayer artificial neural network*. Can Urol Assoc J, 2019. **13**(5): p. E145-e150.
348. Patel, N., A. Henry, and A. Scarsbrook, *The value of MR textural analysis in prostate cancer*. Clin Radiol, 2019. **74**(11): p. 876-885.
349. Hamm, C.A., et al., *Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI*. Eur Radiol, 2019. **29**(7): p. 3338-3347.
350. Dekker, A., et al., *Rapid learning in practice: a lung cancer survival decision support system in routine patient care data*. Radiother Oncol, 2014. **113**(1): p. 47-53.
351. *Cancer imaging archive*. Frederick Nat. Lab for Cancer Research.: United States.

352. Clark, K., et al., *The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository*. Journal of Digital Imaging, 2013. **26**(6): p. 1045-1057.
353. Woźnicki, P., et al., *Multiparametric MRI for Prostate Cancer Characterization: Combined Use of Radiomics Model with PI-RADS and Clinical Parameters*. Cancers (Basel), 2020. **12**(7).
354. Aldoj, N., et al., *Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network*. Eur Radiol, 2020. **30**(2): p. 1243-1253.
355. Min, X., et al., *Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: Cross-validation of a machine learning method*. Eur J Radiol, 2019. **115**: p. 16-21.
356. Quibim, *qp-Prostate*. 2021: Valencia, Spain. p. <https://quibim.com/quibim-receives-fda-510k-clearance-for-qp-prostate-new-ai-solution-for-prostate-mri-analysis/>.
357. Quantib, *Quantib Prostate*. 2021: Rotterdam, Netherlands. p. <https://www.quantib.com/en/solutions/quantib-prostate>.
358. Ezra, *Ezra Prostate AI*. 2020. p. <https://ezra.com/prostate-ai/>.
359. El Naqa, I., et al., *Radiation Therapy Outcomes Models in the Era of Radiomics and Radiogenomics: Uncertainties and Validation*. Int J Radiat Oncol Biol Phys, 2018. **102**(4): p. 1070-1073.
360. Soodana-Prakash, N., et al., *Entering an era of radiogenomics in prostate cancer risk stratification*. Transl Androl Urol, 2018. **7**(Suppl 4): p. S443-s452.
361. Aiello, M., et al., *The Challenges of Diagnostic Imaging in the Era of Big Data*. J Clin Med, 2019. **8**(3).

## 8. LEBENS LAUF

### PERSONALIEN

|                   |                               |
|-------------------|-------------------------------|
| Vorname und Name: | Sarah Elena Schenk            |
| Geburtsdatum:     | 23.09.1993                    |
| Geburtsort:       | Mannheim                      |
| Vater:            | Dr. med. Jochen Walter Schenk |
| Mutter:           | Christiana Maria Schenk       |
| Geschwister:      | Christoph Jochen Schenk       |

### SCHULISCHER WERDEGANG

|             |   |
|-------------|---|
| 2000 – 2004 | Grundschule Lonnig                      |
| 2004 – 2010 | Görres Gymnasium Koblenz                |
| 2010 – 2011 | Rosemount High School, Montreal, Kanada |
| 2011 – 2013 | Gymnasium auf dem Asterstein Koblenz    |
| April 2013  | Erwerb der Allgemeinen Hochschulreife   |

### UNIVERSITÄRER WERDEGANG

|            |   |
|------------|---|
| WS 2014/15 | Beginn des Studiums der Humanmedizin an der Ludwig-Maximilians-Universität München        |
| 09.08.2017 | 1. Abschnitt der Ärztlichen Prüfung   |
| WS 2017/18 | Hauptstudium an der Ruprecht-Karls-Universität Heidelberg, Medizinische Fakultät Mannheim |

|                 |   |
|-----------------|---|
| Seit 06/2018    | Doktorandin an der Klinik für Radiologie und Nuklearmedizin,<br>Universitätsmedizin Mannheim                          |
| 08.10.2020      | 2. Abschnitt der Ärztlichen Prüfung   |
| WS 2020/21      | Praktisches Jahr an der Universität zu Köln   |
| 26.11.2021      | 3. Abschnitt der Ärztlichen Prüfung   |
| Seit 01.04.2022 | Ärztin in Weiterbildung am Institut für Diagnostische und<br>Interventionelle Radiologie am Universitätsklinikum Köln |

## 9. DANKSAGUNG

Ohne die Unterstützung einiger Menschen wäre mir das Anfertigen dieser Dissertation nicht möglich gewesen. Daher ist es mir ein besonderes Anliegen, diesen meine herzliche Dankbarkeit auszudrücken.

An erster Stelle möchte ich mich bei Frau Univ.-Prof. Dr. med. Ulrike Attenberger für ihr Vertrauen und die Überlassung des unglaublich spannenden und aktuellen Themas dieser Arbeit bedanken, sowie für die unkomplizierte Weiterbetreuung meiner Arbeit über mehrere Bundeslandgrenzen hinweg.

Besonderen Dank schulde ich auch Frau Univ.-Prof. Dr. med. Bettina Baeßler, die mich bei meinen anfänglichen Segmentierungen und weit darüber hinaus unterstützte, nicht zuletzt mit geduldigen Zoom-Konferenzen.

Außerdem gilt meine große Dankbarkeit auch Piotr Woźnicki, der mit seinem (nicht zuletzt informatischen) Können diese Arbeit zum Erfolg brachte.

Ich möchte meiner Familie und meinen Freunden danken – für alles. Besonders aber für ein jederzeit offenes Ohr an guten und an weniger guten Doktorarbeits-Tagen, für immer konstruktiven Input und unermüdliche Ermutigung, dafür, immer an mich geglaubt zu haben.

Von ganzem Herzen möchte ich mich bei meinen Eltern bedanken, die mir in meinem Leben alles ermöglicht und mir mit diesem Studium ein Geschenk von unschätzbarem Wert gemacht haben.

Meine Dankbarkeit euch gegenüber und was mir unsere Familie bedeutet, ist nicht in Worte zu fassen.

