Künstliche Intelligenz zur Computer-assistierten Diagnose in PSMA-PET/CT-Untersuchungen bei Prostatakarzinompatienten

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Hohen Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Annette Erle

aus Paderborn

2023

Angefertigt mit der Genehmigung

der Medizinischen Fakultät der Universität Bonn

- 1. Gutachter: Prof. Dr. Dr. Ralph A. Bundschuh
- 2. Gutachter: Prof. Dr. Peter Michael Krawitz

Tag der Mündlichen Prüfung: 20.01.2023

Aus der Klinik und Poliklinik für Nuklearmedizin Direktor*in: Prof. Dr. Markus Essler

Inhaltsverzeichnis

	Abkürzungsverzeichnis	4
1.	Deutsche Zusammenfassung	6
1.1	Einleitung	6
1.2	Material und Methoden	12
1.3	Ergebnisse	16
1.4	Diskussion	17
1.5	Zusammenfassung	20
1.6	Literaturverzeichnis der deutschen Zusammenfassung	21
2.	Veröffentlichung	27
	Abstract	27
	Introduction	27
	Materials and Methods	28
	Results	31
	Discussion	35
	References	37
3.	Danksagung	39

Abkürzungsverzeichnis

AUC	area under the curve
BiTEs	bispecific T-cell engagers
BRPC	biochemically recurring prostate cancer
CD3	cluster of differentiation
СТ	Computertomographie
CV	cross validation
DW-MRI	diffusion-weighted magnetic resonance imaging
GI	Glandula
mCRPC	metastatic castration resistant prostate cancer
Met	Metastase
ML	machine learning
ml	Milliliter
MRT	Magnetresonanztomographie
ng	Nanogramm
PET	Positronen-Emissions-Tomographie
PSA	Prostataspezifisches Antigen
PSMA	Prostataspezifisches Membranantigen
ROI	region of interest
SPECT	single photon emission computed tomography
SRT	salvage radiotherapy
SUV	standardized uptake value
SVM	support vector machine
VOI	volume of interest
WHO	World Health Organization
¹⁸ F	Fluor-18
¹⁸ F-PSMA-1007	(((3 <i>S</i> ,10 <i>S</i> ,14 <i>S</i>)-1-(4-(((<i>S</i>)-4-carboxy-2-((<i>S</i>)-4-carboxy-2-(6- ¹⁸ F-
	fluoronicotinamido)butanamido)butanamido)methyl)phenyl)-
	3-(naphthalen-2-ylmethyl)-1,4,12-trioxo-2,5,11,13-
	tetraazahexadecane-10,14,16-tricarboxylic acid))

2-(3-{1-carboxy-5-[(6-18F-fluoro-pyridine-3-carbonyl)-
amino]-pentyl}-ureido)-pentanedioic acid
Gallium-68
Lutetium-177
Actinium-225

1. Deutsche Zusammenfassung

1.1 Einleitung

Maschinelles Lernen (ML) und künstliche Intelligenz spielen in vielen Bereichen eine immer größere Rolle. Auch in der Medizin werden solche Ansätze zunehmend integriert und erforscht, was die steigende Anzahl an Publikationen zu diesem Themenfeld zeigt (National Library of Medicine, 2021). Der demographische Wandel und die damit einhergehende Zunahme der benötigten Kapazitäten im Gesundheitssystem machen die Anwendung dieser Methoden umso interessanter und notwendiger. Aufgaben, die heute von einem Arzt bearbeitet werden müssen, können noch künftig von Computerprogrammen unterstützt und in Zukunft sogar von ihnen übernommen werden. Dies erlaubt dem medizinischen Personal sich verstärkt anderen Tätigkeiten zu widmen, die tatsächlich menschliche Qualitäten erfordern, wie beispielsweise die Aufklärung des Patienten und die individuelle Beratung zu Therapieoptionen.

Besonders in der medizinischen Bildgebung ergeben sich diesbezüglich viele Möglichkeiten. So kann beispielsweise die Analyse von Computertomographie (CT) -Bildern anhand von maschinellem Lernen möglicherweise effizienter, kostengünstiger, und insbesondere schneller erfolgen. Die Menge an Daten nimmt besonders in diesem Gebiet zu, da die Untersuchungen heute flächendeckend schnell verfügbar sind und deshalb häufiger durchgeführt werden. Außerdem erlauben auch der technische Fortschritt und die hiermit verbundenen neuen Möglichkeiten noch genauere Untersuchungsmöglichkeiten, wie zum Beispiel die Umsetzung einer Schnittbildgebung mit noch dünneren Schichten, was wiederum in einem Anstieg an zu verarbeitenden und auszuwertenden Informationen resultiert

Ziel dieser Arbeit ist es, die Anwendung von maschinellem Lernen auf die PSMA (Prostata-spezifisches Membranantigen)-Bildgebung bei Prostatakarzinom zu untersuchen. Im Folgenden sollen sowohl die Epidemiologie des Prostatakarzinoms als auch themenrelevante Literatur zur Bedeutung des PSMA und Grundlagen des maschinellen Lernens einführend dargelegt werden, um dann zur Fragestellung dieser Arbeit hinzuleiten.

1.1.1 Epidemiologie

Der World Health Organization (WHO) zufolge steht das Prostatakarzinom an Platz zwei der häufigsten malignen Neoplasien bei Männern weltweit, mit insgesamt 1 414 259 Fällen im Jahr 2020 (WHO-International Agency for Research on Cancer, 2021). Bis zum Jahr 2040 ist in der Gruppe der Männer über 70 Jahre mit einem Zuwachs der Inzidenz um 89,1 % zu rechnen, auch die Mortalität wächst um 106 % (WHO-International Agency for Research on Cancer, 2021). Diese Zahlen verdeutlichen die aktuelle, sowie auch die zukünftige medizinische Relevanz der Erkrankung, insbesondere vor dem Hintergrund des demographischen Wandels mit einer Zunahme der älteren Bevölkerung. Die relative 5-Jahres-Überlebensrate bei Prostatakarzinom liegt insgesamt bei 97,5 % (National Cancer Institute, 2021). Handelt es sich jedoch um ein fortgeschrittenes Stadium mit Fernmetastasen, sinkt sie auf 30,6 % (National Cancer Institute, 2021). Dies macht deutlich, wie wichtig ein verlässliches *Staging* am Anfang der Diagnostik für die Prognose ist, um mögliche Metastasen nicht zu übersehen und so eine stadiengerechte Therapie zu gewährleisten.

1.1.2 PSMA

Das Prostata-spezifische-Membran-Antigen N-Acetyl-L-aspartyl-L-(auch glutamatpeptidase I) ist ein aus 750 Aminosäuren bestehendes Transmembranprotein, was als Glutamat-Carboxypeptidase fungiert (Heck et al., 2017; Jones et al., 2020). In gesunden Zellen befindet es sich im Zytosol und am apikalen Pol der Zelle (Heck et al., 2017). Unter der malignen Transformation wird es an der luminalen Membran exprimiert (Jones et al., 2020). Die Funktion des Proteins bleibt bis auf weiteres unklar (Jones et al., 2020). Möglicherweise übernimmt es Transportfunktionen im Rahmen der Endozytose (Jones et al., 2020). Obwohl verschiedene Studien gezeigt haben, dass PSMA in unterschiedlichen Geweben (neben der Prostata auch in Speicheldrüsen, Duodenum, Niere, Kolon) in niedriger Konzentration exprimiert wird und es somit nicht spezifisch für das Prostatakarzinom ist, zeigen sich höhere Expressionen des Proteins an typischen Lokalisationen der Metastasierung wie Lymphknoten und Knochen nur bei Prostatakarzinom (Heck et al., 2017; Jones et al., 2020; Silver et al., 1997). Immunhistochemisch ist ein erhöhter PSMA-Wert insbesondere mit dedifferenzierten,

metastasierten oder hormon-refraktären Tumoren assoziiert und korreliert außerdem mit der Prognose des Patienten (García Garzón et al., 2018).

Strukturen, die an PSMA binden, haben heute Bedeutung sowohl in der Diagnostik als auch in der Therapie.

1.1.2.1 PSMA in der Bildgebung

Radioaktiv markierte PSMA-Liganden, die mittels nuklearmedizinischer Diagnostik sichtbar gemacht werden können, sind heute etablierte Mittel im Staging des Prostatakarzinoms. Erste Ansätze wurden mit Antikörper-basierten Technologien durchgeführt, wobei diese den heute etablierten niedermolekulare PSMA-Liganden durch höhere Strahlenbelastung und deutlich niedrigere Penetration des Tumorgewebes unterlegen sind (Jones et al., 2020). Insgesamt ist der Ansatz auf SPECT (single photon emission computed tomography), PET/MRT (Hybridverfahren aus Positronen-Emissions-Tomographie und Magnetresonanztomographie) und PET/CT (Hybridverfahren aus Positronen-Emissions-Tomographie und Computertomographie) anwendbar, wobei letzteres den breitesten klinischen Einsatz erfährt. Der heute in der Diagnostik meist genutzte Ligand ist ⁶⁸Ga-PSMA-11 (Jones et al., 2020). Im PET/CT wird die dreidimensionale Verteilung des mit Gallium-68 markierten PSMA-Liganden gemessen, welches sich in maligne transformiertem Gewebe vermehrt anreichert (García Garzón et al., 2018). Die Kombination von PET und CT, was als Hybridbildgebung bezeichnet wird, ermöglicht eine genaue anatomische Zuordnung der nuklearmedizinischen Funktionsdiagnostik und wird besonders bei onkologischen Fragestellungen hinzugezogen (Kahl-Scholz und Vockelmann, 2017).

Studien zeigen die Überlegenheit der PSMA-basierten Bildgebung im Vergleich zur konventionellen Bildgebung (Jones et al., 2020). Tan et al. demonstrieren in ihrer systematischen, 43 Studien umfassenden Übersichtsarbeit, dass PSMA-markierte *radiotracer* besonders bei niedrigen Werten des Prostataspezifischen Antigens (PSA<0,5 ng/ml) und BRPC (*biochemically recurrent prostate cancer* nach totaler Prostatektomie) hohe Detektionsraten von 44,9 % erzielen (Tan et al., 2019).

Bei der Detektion von Lymphknotenmetastasen ist die PSMA-basierte Bildgebung der konventionellen Bildgebung (MRT und *diffusion-weighted magnetic resonance imaging*

(DW-MRI)) in Sensitivität und Spezifität überlegen (Petersen et al., 2020). Zhou et al. zeigen in ihrer Metaanalyse, dass auch bei der Detektion von Knochenmetastasen die PSMA-PET/CT konventionellen Verfahren, wie der Knochenszintigraphie und dem MRT überlegen ist (Zhou et al., 2019). Auch im Vergleich zur *low dose* CT weist sie eine höhere Sensitivität auf (Sachpekidis et al., 2018). In einer weiteren Studie wurde bei zuvor kurativ behandelten Patienten mit nun biochemischer Rekurrenz zunächst eine PET/CT-Untersuchung mit Flour-18-Cholin (¹⁸F-Cholin) durchgeführt und wenn dieses negativ war eine ⁶⁸Ga-PSMA-PET/CT angeschlossen. Die ⁶⁸Ga-PSMA-PET/CT konnte in 43,8 % den Ursprung des biochemischen Rezidivs identifizieren (Bluemel et al., 2016).

Neue PSMA-Radioliganden, wie beispielsweise ¹⁸F-PSMA-1007 und ¹⁸F-DCFPyL, zeigen eine vergleichbare Sensitivität bei BRPC Patienten, aber auch mögliche Vorteile, wie eine bessere Bildqualität mit deutlicherer Sichtbarkeit von kleinen Metastasen, eine geringere Anreicherung im Harntrakt und dadurch eine vereinfachte Beurteilung des Beckenbereichs (Giesel et al., 2019; Kuten et al., 2020). Weiterhin verfügen die neuen Liganden über eine längere Halbwertszeit, was Transport und Produktion im Gegensatz zu Gallium-68 vereinfacht (Jones et al., 2020).

In verschiedenen Studien demonstrieren die neuen Fluor-basierten-Liganden ein genauso hohes diagnostisches Potenzial ohne signifikanten Unterschied im *Staging* und *Re-Staging* im Vergleich zu ⁶⁸Ga-PSMA mit akkurater Identifizierung der Läsionen (Lin et al., 2019). Giesel et al. schlussfolgern eine vergleichbare oder sogar bessere Detektionsrate von ¹⁸F-PSMA-1007 im Vergleich zu ⁶⁸Ga-PSMA-Liganden (Giesel et al., 2019). Ferner zeigt der Ligand ¹⁸F-DCFPyL bei PSA-Werten unter 2,0 ng/ml eine bessere Entdeckung von Metastasen als ⁶⁸Ga-PSMA (Wondergem et al., 2019).

Klinische Anwendungsmöglichkeiten der ligandenbasierten Bildgebung finden sich zum Beispiel im *Re-Staging* bei BRPC. Außerdem sorgt besonders die hohe Sensitivität bei niedrigen PSA-Werten für eine gezieltere Therapieeinleitung beispielsweise in die SRT (*salvalge radiotherapy*). Aufgrund der oben beschriebenen Eigenschaften eignet sich die PSMA-PET/CT auch als präoperatives Lymphknoten-*Staging*, zur Planung einer gezielten Biopsieentnahme und zur Identifizierung von Patienten, die für eine PSMA-Radiotherapie in Frage kommen (Jones et al., 2020).

9

1.1.2.2 **PSMA** in der Endoradiotherapie

Auch in der Therapie wird versucht, die Tumorzellen gezielt auszuschalten, beispielsweise mittels des mit einem PSMA-Inhibitor verbundenen Betastrahlers Lutetium-177 (¹⁷⁷Lu-PSMA-617) (Rahbar et al., 2018). Somit stellt PSMA einerseits eine vielversprechende Struktur für Bildgebung und Diagnostik und andererseits für die gezielte Therapie, die sogenannte *targeted therapy* dar.

Die Radionuklidtherapie kommt in sehr fortgeschrittenen Krankheitsstadien zum Einsatz (bei metastasiertem kastrationsresistentem Prostatakarzinom) (AWMF, Deutsche Krebsgesellschaft, Deutsche Krebshilfe, 2021). Der PSMA-Ligand wird mit einem Alphaoder Betastrahler markiert, was die Tumorzelle letztlich zerstört. Heute werden bevorzugt niedermolekulare PSMA-Inhibitoren eingesetzt, da diese weniger hämatotoxisch wirken als Antikörper-basierte Therapien (Jones et al., 2020). Betastrahler zeichnen sich gegenüber Alphastrahlern durch eine längere Halbwertszeit und eine größere Reichweite der radioaktiven Strahlung mit Schonung der gesunden Zellen aus (Jones et al., 2020). Am häufigsten genutzt werden die Betastrahler ¹⁷⁷Lu-PSMA-I&T und ¹⁷⁷Lu-PSMA-617, wobei letzteres das bessere pharmakokinetische Profil aufweist (Jones et al., 2020). Alpha-Strahler wie ²²⁵AC-PSMA-617 können zum Beispiel bei Infiltration des roten Knochenmarks und Resistenz gegen Betastrahler eingesetzt werden (Kratochwil et al., 2016).

Voraussetzung für die Radiotherapie ist der Nachweis einer Ligandenbindung mit PSMA-Expression der Läsionen in der PSMA-PET (Heck et al., 2017). Kreish et al. evaluieren in einer prospektiven Studie mit 254 Männern mit metastasiertem kastrationsresistentem Prostatakarzinom (mCRPC) die Therapie mit ¹⁷⁷Lu-PSMA-617. Sie bewerten die Therapie als effektiv, sicher und gut verträglich (Khreish et al., 2021). Die häufigsten Nebenwirkungen waren Anämie, Thrombozytopenie und Lymphopenie sowie eine durch die Radiotherapie bedingte Xerostomie (Khreish et al., 2021). Auch eine retrospektive Analyse von 248 Therapiezyklen mit ¹⁷⁷Lu-PSMA-617 belegt deren Sicherheit und stellt außerdem die überlegene Wirksamkeit gegenüber anderen Dritt-Linien-Therapien fest (Rahbar et al., 2017). Das Gesamtüberleben zeigt sich vergleichbar mit anderen Dritt-Linien-Therapien (Jones et al., 2020). Ein neuer Ansatz ist der Einsatz von BiTEs *(bispecific T cell engagers)*. Diese neuartigen Antikörper wirken immunmodulatorisch und induzieren die antigenspezifische Aktivierung des Immunsystems (Kamat et al., 2021). BiTEs verbinden zytotoxische T-Zellen über CD3 *(cluster of differentiation)* mit den PSMA-positiven Tumorzellen (Kamat et al., 2021). Die zytotoxischen T-Zellen induzieren daraufhin über Perforine und Granzym die Apoptose der malignen Zellen (Hummel et al., 2019). Verschiedene BiTEs wie AMG212 und AMG160 zeigen ihre signifikante Wirkung bei mCRPC und sind Gegenstand aktueller Forschung (Hummel et al., 2021; Kamat et al., 2021).

1.1.3 Maschinelles Lernen

Maschinelles Lernen generiert Wissen aus vorhandenen Daten mit dem Ziel, die Daten zu klassifizieren und Vorhersagen zu treffen. In diesem Forschungsfeld finden sich Elemente aus der Statistik, der künstlichen Intelligenz und der Informatik (Müller und Guido, 2017).

In der Wissenschaft, aber auch in vielen anderen Bereichen sind die Datenmengen heute zu groß, als dass sie von Fachkräften analysiert und ausgewertet werden könnten (Alpaydin, 2019). Dies erfordert einerseits deutlich mehr zeitliche und finanzielle Ressourcen, andererseits eine Vielzahl an Arbeitskräften. Daher bietet sich die Anwendung von ML als Lösung an.

Im Grundsatz werden beim ML Schlussfolgerungen aufgrund von Stichproben gezogen. Diese Stichproben ergeben einen Trainingsdatensatz, worauf ein Algorithmus ein statistisches Modell aufbaut (Alpaydin, 2019). Muster und Regelmäßigkeiten in den Daten werden so erkannt, mit dem Ziel, diese auch auf neue unbekannte Daten übertragen zu können und so möglichst akkurate Vorhersagen zu treffen (Alpaydin, 2019).

Man unterscheidet das überwachte Lernen *(supervised learning)* einerseits vom unüberwachten Lernen *(unsupervised learning)* andererseits. Beim überwachten Lernen, was auch im Fall dieser Arbeit zur Anwendung kommt, lernt der Algorithmus anhand von Eingabe-Ausgabe-Paaren (Müller und Guido, 2017). Die korrekte Ausgabe ist hier bekannt und es sollen Assoziationen hergestellt werden (Alpaydin, 2019; Müller und Guido, 2017). Beim unüberwachten Lernen geht es um die Erkennung von Mustern allein in den Eingabewerten, die Ausgabe ist hier nicht bekannt (Alpaydin, 2019). Mit *Clustering*-

Verfahren, also der Analyse von Häufungen oder Gruppierungen, können die Eingabewerte sortiert werden (Alpaydin, 2019).

Zum Einsatz kommen ML-Ansätze heute in sehr verschiedenen Bereichen, wie zum Beispiel in der Medizin, bei der SPAM-Erkennung von E-Mails, in der Spracherkennung, im Finanzwesen, in der Telekommunikation, in der Erstellung von personalisierter Werbung und in den Naturwissenschaften mit dem Ziel der Wissensextraktion und - kompression (Alpaydin, 2019).

1.1.4 Fragestellung der Arbeit

Ich beschäftigte mich in dieser Arbeit mit der Frage, inwieweit ein Algorithmus in der Lage ist, mithilfe von maschinellem Lernen die pathologische Aufnahme des Markierungsstoffes ⁶⁸Ga-PSMA in der PET/CT zu erkennen.

Die Studie untersucht und vergleicht dazu die Anwendung von drei Algorithmen basierend auf überwachtem maschinellem Lernen auf die PSMA-PET/CT-Daten von 72 Patienten mit Prostatakarzinom. Das Programm klassifiziert die sogenannten *hotspots* in die Kategorien "physiologisch" und "pathologisch". Als *hotspot* wird in der Nuklearmedizin eine Stelle mit deutlich erhöhter Radionuklidbelegung verstanden. Um die Treffsicherheit unseres Algorithmus anschließend zu validieren, testen wir die Genauigkeit der Klassifikation mit den PET/CT-Bildern einer unabhängigen Kohorte aus 15 weiteren Patienten.

1.2 Material und Methoden

In dieser Studie werden die Daten von 87 Patienten untersucht, bei denen histologisch ein Prostatakarzinom nachgewiesen war. Alle erhielten eine PSMA-PET/CT mit ⁶⁸Ga-HBED-CC-PSMA (⁶⁸Ga-PSMA-11). Zunächst wurde eine *low dose* CT von Schädel bis Mitte Oberschenkel angefertigt. Dann folgte eine PET des gleichen Abschnitts mit 3 oder 4 Minuten pro Bettposition abhängig vom Gewicht des Patienten.

Die PSMA-PET/CT-Scans von 72 Patienten dienten als Trainingsset. Alter, Gleason-Score und PSA-Wert wurden erfasst und sind in Table 1 der Publikation genau aufgelistet (Erle et al., 2021, Table 1). Die Bilddaten von 42 Patienten analysierte ich mithilfe der InterView Fusion Software von Mediso Medical Imaging. Die Daten der verbleibenden 30 Patienten wurden bereits in einer vorherigen Studie auf die gleiche Weise untersucht und lagen mir vor. Ich markierte die Aufnahme des Markierungsstoffes als VOI (volume of interest) manuell mithilfe eines Pinselwerkzeugs (brushtool). Dieser Prozess umschloss jegliche Aufnahme des Markierungsstoffes ⁶⁸Ga-PSMA ohne vorher festgelegten SUV (standardized uptake value), zum Beispiel den uptake in Drüsen, Niere aber auch Knochen und Lymphknoten. Die Markierung erfolgte Schnittbild-pro-Schnittbild, das heißt, auf jedem Schnittbild wurden die Läsionen erneut markiert, woraus sich über eine dreidimensionale Rekonstruierung ein Volumen (VOI) ergab (siehe Abbildung 1). Aufgrund der anatomischen Lokalisation, die durch die Kombination der PET mit der CT klar zuzuordnen war, teilte ich anschließend die markierten Läsionen in die Kategorien "physiologisch" und "pathologisch" ein. Beispielsweise ist sowohl die Aufnahme in Drüsen, wie auch in Nieren und Harnblase als physiologisch im Sinne einer Konzentrierung und Ausscheidung im Urin beziehungsweise Speichel zu werten. Anreicherungen in knöchernen Strukturen und Lymphknoten dagegen wurden der Kategorie "Met" (für Metastase = pathologisch) zugeteilt (siehe Abbildung 2). Insgesamt ergab sich so eine Zahl von 2452 hotspots mit 1634 pathologischen und 818 physiologischen ⁶⁸Ga-PSMA-Anreicherungen. Um die Spezifität der Vorhersage später genauer beurteilen zu können, teilte ich die hotspots zum einen in die Kategorie "Met" für jeglichen pathologischen tracer uptake ein, zum anderen gliederte ich die Kategorie "physiologisch" in weitere Subgruppen (siehe Tabelle 1).

Abb. 1: Beispiel-Screenshot der Bearbeitung mit der InterView Fusion Software. Links PSMA-PET/CT, rechts *low-dose* CT.

Die Anreicherung von ⁶⁸Ga-PSMA wurde Schnittbild für Schnittbild markiert, wodurch sich nach dreidimensionaler Rekonstruktion VOI ergaben.



Im Anschluss wählte ich 15 weitere Patienten aus, die in Alter, Gleason Score und PSA-Wert unserer Trainingskohorte ähnelten (Erle et al., 2021, Table 1). Diese bildeten die Testkohorte. In den Scans dieser 15 Patienten markierte ich pro Patient 5 bis 10 pathologische Läsionen, sowie jegliche Aufnahme von ⁶⁸Ga-PSMA in Drüsen und Leber, außerdem 5 weitere physiologische *hotspots*. Hier resultierten 331 *hotspots* mit 128 pathologischen und 203 physiologischen Läsionen.

Eine Untergruppe von 30 Patienten wurde als erste Trainingsgruppe genutzt, um die entwickelte Software in Python V.3.5 einzustellen und mit anderen ML *classifiers* zu vergleichen. Die Entwicklung der Software, die anschließende Einstellung und das Training wurden von Sobhan Moazemi als Informatiker in Zusammenarbeit mit der mathematisch-naturwissenschaftlichen Fakultät der Universität Bonn durchgeführt. Da durch meine vorherige Einteilung in die Kategorien "physiologisch"/"pathologisch" sowohl Eingabe als auch Ausgabewerte vorhanden waren, handelt es sich hier um einen Fall des überwachten Lernens. Evaluiert wurden drei *machine learning classifiers* der SciKitLearn library mittels *cross-validation (CV)*, darunter lineare *support vector machine (SVM), Extra Trees* und *random forest*. Bewertet wurden die Algorithmen nach *AUC (area under the curve)*, Sensitivität und Spezifität mit Standardabweichung und Mittelwert. Hier stellte sich im Vergleich *Extra Trees* als der *classifier* heraus, der die präzisesten Ergebnisse lieferte. Daher verwendeten wir *Extra Trees* für den folgenden Trainingsprozess und die anschließende Testung der Anwendbarkeit auf unbekannte Daten mittels der Testkohorte.

Abb. 2: Beispiel-Tabelle "ROI Statistics" mit Aufführung der verschiedenen hotspots. Die Läsionen wurden anhand ihrer anatomischen Lokalisation eingeteilt und beschriftet.

1	ROI Statistics				- 🗆	×
	A (Orientation)	B (Name)	C (Color)	D (Image name)	E (Deviation)	F
2	[Ax]	Met12		PET(g/ml)	1.37	2.9
3	[Ax]	Met8		PET(g/ml)	1.15	2.3
4	[Ax]	Met9		PET(g/ml)	3.59	4.7
5	[Ax]	Met10		PET(g/ml)	2.22	5.3
6	[Ax]	Niere re		PET(g/ml)	11.83	25.
7	[Ax]	Met13		PET(g/ml)	2.15	2.8
8	[Ax]	Blase		PET(g/ml)	2.50	9.7
9	[Ax]	Met15		PET(g/ml)	1.98	3.9
10	[Ax]	Met18		PET(g/ml)	1.03	2.5
11	[Ax]	Met7		PET(g/ml)	0.9326	2.4
12	[Ax]	Met5		PET(g/ml)	1.27	1.9
13	[Ax]	Met16		PET(g/ml)	2.44	6.4
14	[Ax]	Met6		PET(g/ml)	0.9807	3.9
15	[Ax]	Met4		PET(g/ml)	1.27	2.0
16	[Ax]	Met11		PET(g/ml)	1.02	3.2
17	[Ax]	Niere li		PET(g/ml)	11.78	24.
18	[Ax]	Met14		PET(g/ml)	1.86	3.4
19	[Ax]	Met17		PET(g/ml)	0.9780	2.2
20	[Ax]	Darm		PET(g/ml)	1.78	3.5
21	[Ax]	Met20		PET(g/ml)	0.7024	1.5
22	[Ax]	Met19		PET(g/ml)	0.9120	1.5
23	[Ax]	Met21		PET(g/ml)	0.9870	1.7
24	[Ax]	Met22		PET(g/ml)	0.7684	2.2
25	[Ax]	Met23		PET(g/ml)	0.5909	1.6
26	[Ax]	Met24		PET(g/ml)	0.6361	2.4
27	[Ax]	Parotis re		PET(g/ml)	2.33	6.2
28	[Ax]	Submandibularis re	e 💶	PET(g/ml)	2.68	6.8
29	[Ax]	Met1		PET(g/ml)	3.98	7.5
30	[Ax]	Met2		PET(g/ml)	2.40	4.9
31	[Ax]	Submandibularis li		PET(g/ml)	2.34	8.0
32	[Ax]	Met3		PET(g/ml)	1.19	4.6
33	[Ax]	Lacrimalis re		PET(g/ml)	2.15	5.3
34	[Ax]	Lacrimalis li		PET(g/ml)	1.19	3.7
35	[Ax]	Sublingualis re		PET(g/ml)	0.8086	2.9
36	[Ax]	Sublingualis li		PET(g/ml)	1.17	4.1
37	[Ax]	Parotis li		PET(g/ml)	1.64	6.2
•						Þ
	2	F	+	×	Current layer	•

Kategorie	pathologisch	physiologisch
Untergruppen	Met	Niere
		Ureter
		Blase
		Darm
		Glandula lacrimalis
		Glandula parotis
		Glandula sublingualis
		Glandula submandibularis
		Prostata
		Leber

Tab. 1: Einteilung der Kategorien physiologisch und pathologisch mit Untergruppen

1.3 Ergebnisse

Insgesamt ergaben sich 2452 *hotspots* in der Trainingskohorte mit 818 als physiologisch und 1634 als pathologisch klassifizierten Läsionen. In der Kontrollgruppewaren waren es 331 *hotspots* insgesamt, inklusive 128 pathologischen und 203 physiologischen Anreicherungen von ⁶⁸Ga-PSMA (Erle et al., 2021, Table 3 und 4).

Der *Extra Trees classifier* war den Algorithmen *SVM* und *random forest* überlegen, was sich während der *cross-validation* zeigte (Erle et al., 2021, Figure 2). Nach dem Training mit den Daten von 72 Patienten erreichte er eine maximalen AUC von 0,98, eine Sensitivität von 0,97 und eine Spezifität von 0,83. (Erle et al., 2021, Table 5). 125 von 128 in der Testgruppe als "Met" eingestufte Läsionen wurden vom Algorithmus als pathologisch erkannt, was in einer Sensitivität von 0,97 resultiert. Auch bezüglich der physiologischen *hotspots* konnte eine gute Prädiktion erreicht werden. Hier zeigten sich hervorragende Ergebnisse für Niere, Darm und Leber mit 0 falsch-positiven Zuordnungen (Spezifität = 1,00), wohingegen die Rate der falsch-positiven Einteilungen bei den Drüsen mit 19 von 111 deutlich höher war (Spezifität = 0,82). Besonders in den sublingualen und lacrimalen Drüsen zeigten sich Schwierigkeiten (Glandula lacrimalis 7 von 28 falsch positiv, Spezifität = 0,75; Glandula sublingualis 9 von 23 falsch positiv, Spezifität = 0,61)

(Erle et al., 2021, Table 6). Die genauen Ergebnisse aufgeschlüsselt nach Kategorie mit Untergruppen sind in Tabelle 2 aufgeführt.

Kategorie	1/pathologisch	0/physiologisch	Total	Spezifität
Blase	0	13	13	1,00
Drüsen	19	92	111	0,83
Gl sub-	2	28	30	0,93
Gl sublingualis	9	14	23	0,61
GI lacrimalis	7	21	28	0,75
GI parotis	1	29	30	0,97
Darm	1	32	33	0,97
Leber	0	15	15	1,00
Nieren	0	30	30	1,00
Ureter	0	1	1	1,00
Metastasen	125	3	128	

Tab. 2: Ergebnisse der Prädiktion in der Testkohorte

1.4 Diskussion

Die Anwendungs- und Einsatzmöglichkeiten von ML im medizinischen Kontext nehmen zu (Handelman et al., 2018). Besonders in der medizinischen Bildgebung ist die Implementation vielversprechend, da verschiedene Schritte im Arbeitsablauf erleichtert, verbessert oder sogar komplett übernommen werden können (Choy et al., 2018). In den letzten Jahren wurde dies auch zunehmend Gegenstand der Forschung. So zeigen sich bereits in unterschiedlichen Bereichen aussichtsreiche Ergebnisse, wie beispielsweise im Mamma-MRT, in der Beurteilung von Lungenrundherden oder in der Analyse des Knochenalters in der pädiatrischen Radiologie und Endokrinologie (Choy et al., 2018; Reig et al., 2020). Um eine zukunftsweisende Medizin zu gestalten, die präventiv und prädiktiv arbeitet und überdies individualisierte Therapien ermöglicht, ist es unumgänglich, die Digitalisierung auch im medizinischen Bereich voranzutreiben und die Analyse von großen Datenmengen (*big data*) einzusetzen (Alonso-Betanzos und Bolón-Canedo, 2018). Aufgrund des zu erwartenden Anstiegs der Inzidenz und Prävalenz des Prostatakarzinoms ist die Akquirierung großer Datenmengen unkompliziert, daher ist die Erkrankung als Studienobjekt optimal. Auch in der praktischen Umsetzung würde eine

Vielzahl von Patienten profitieren. Bedingt durch die große Anzahl der PSMA-PET/CT-Untersuchungen wäre zudem die Erleichterung der Arbeitsabläufe in der Klinik beachtlich.

Trotz allem muss kritisch diskutiert werden, wie und in welchem Umfang ML die heutigen Abläufe unterstützen und/oder ersetzten kann und darf. Hier sind verschiedene Überlegungen zu berücksichtigen, wie zum Beispiel der Umgang mit personenbezogenen Daten. Der klinische Alltag erfordert Modelle, die eine maximale Präzision zeigen, wofür die Auswertung von *big data* nötig ist. Daher besteht der permanente Bedarf nach immer größeren Datenmengen. Der Datenschutz und damit auch der Schutz des Patienten vor Missbrauch der Informationen durch Dritte müssen trotzdem jederzeit gewährleistet sein. Weiterhin ist eine zunehmende Abhängigkeit von den computer-technologischen Möglichkeiten zu diskutieren. Ein davon unabhängiges, grundsätzlich eigenständiges Arbeitskonzept sollte erhalten bleiben, um auch in Notfällen, wie zum Beispiel einem Stromausfall, weiterhin handlungsfähig zu sein und eine adäquate Patientenversorgung gewährleisten zu können.

Des Weiteren sollten ethische Fragestellungen diskutiert werden. Dies zeigt sich besonders im Bereich der prädiktiven Möglichkeiten: Hier stellt sich die Frage, in wieweit der einzelne Patient über seine Prognose informiert werden sollte und wie sich dies kommunizieren lässt. Wenn Ärzte nicht mehr genau nachvollziehen können, was in der *black box* des Algorithmus vorgeht, das heißt wie genau dieser zu seinen Schlussfolgerungen kommt, muss man sich damit auseinandersetzten, wie man es ethisch vertreten kann, diese Methoden überhaupt anzuwenden (Anand Shroff, 2017).

Ein Problem bei ML ist auch, dass alle Voraussagen auf den Daten beruhen, mit denen die Software lernt. Gibt es ein Bias in den Daten, ist auch das gesamte darauf aufbauende statistische Modell hinfällig (Anand Shroff, 2017). Letztlich trifft der Algorithmus nur Vorhersagen aufgrund von Korrelationen und lässt nur bedingt Rückschlüsse auf Kausalitäten zu (Obermeyer und Emanuel, 2016). ML bleibt damit ein unterstützendes Werkzeug in der Medizin und ist nicht die Lösung aller gegenwärtigen Probleme in der Diagnostik (Anand Shroff, 2017).

Bezüglich des Ansatzes der ML-basierten Klassifikation von Läsionen in der PSMA-PET/CT bei Prostatakarzinom, der in dieser Arbeit untersucht wird, sind verschiedene Vorteile zu diskutieren. Die PSMA-PET/CT kommt sowohl in der Rezidivdiagnostik bei BRPC, als auch primär bei *high risk* Patienten sowie vor Einleitung einer Therapie mit ¹⁷⁷Lu-PSMA zum Einsatz (AWMF, Deutsche Krebsgesellschaft, Deutsche Krebshilfe, 2021). Da die Untersuchungen ohnehin Bestandteil des diagnostisch-therapeutischen Ablaufs sind, sind keine zusätzlichen Maßnahmen notwendig, was ansonsten für Patient und Gesundheitssystem weiteren zeitlichen als auch finanziellen Aufwand bedeuten würde und zudem mit einer zusätzlichen Strahlenbelastung verbunden wäre. Die Analyse des Bildmaterials mithilfe der Software ermöglicht eine schnelle vorläufige Sortierung, auf deren Grundlage die Kontrolle durch einen Nuklearmediziner denkbar ist. Besonders die hohe Sensitivität mit der der Algorithmus Metastasen erkennt, gewährleistet die erforderliche Sicherheit für den Patienten. Außerdem bietet die dargestellte Vorgehensweise die Möglichkeit, das Therapieansprechen (Moazemi et al., 2021a) sowie das Gesamtüberleben vorherzusagen (Moazemi et al., 2021b).

Die Zusammenarbeit zwischen Arzt und Maschine erlaubt eine zeitsparende Bearbeitung der Bilddaten und so eine Kostenreduzierung bei erhaltener Präzision der Leistung (Deo, 2015). Diese Kooperation gestattet gegebenenfalls eine schnellere Einbettung in die klinische Praxis, da die vollständige Übertragung von Aufgaben an Maschinen gerade im Gesundheitssystem weiter kritisch betrachtet wird (Deo, 2015).

Verbesserungsmöglichkeiten liegen besonders in der Klassifizierung der Speicheldrüsen, da hier die Spezifität bei nur 0,82 liegt. Probleme könnten im Zusammenhang mit einem weit fortgeschrittenen metastasierten Prostatakarzinom auftreten, wenn Metastasen im knöchernen Schädel nicht von submandibularen und lacrimalen Drüsen unterschieden werden können. Die Einteilung des *uptake* in Niere, Blase und Darm erreicht ein deutlich überzeugenderes Ergebnis (mit Spezifitäten von 1,00 und 0,97).

Um Spezifität und Sensitivität insgesamt zu steigern, wäre ein noch umfangreicheres Trainingsset ideal. Da es sich bei dieser Studie um eine retrospektive Analyse handelt, in der bereits vorhandene Bilddaten bearbeitet wurden, befinden sich alle Patienten in einem schon recht fortgeschrittenen Krankheitsstadium. Interessant wäre zusätzlich die Analyse von PSMA-PET/CTs bei Patienten in frühen, nicht metastasierten Stadien, um so die Variabilität in den Eingabewerten zu erhöhen und damit eine noch höhere Vorhersagevalidität zu erreichen. Auch die Analyse von Läsionen in der Prostata selbst ist ein interessanter Ausgangspunkt für weitere Forschungen. In der vorliegenden Untersuchung wurde eine Sensitivität von 0,92 erreicht, die jedoch aufgrund mangelnder Läsionen in der Testkohorte nicht überprüft werden konnten.

Die manuelle Markierung und Sortierung der Läsionen erfordert viel Zeit und sollte in Zukunft automatisch oder halb-automatisch erfolgen. Der Fokus dieser Arbeit lag auf der Evaluation des Algorithmus und somit bleibt die Entwicklung und Evaluierung automatischer Erkennungssysteme Thema weiterführender Studien.

Eine weitere Einschränkung stellt die bei der Anzahl der Läsionen (n~2500) nicht durchführbare bioptisch-histologische Beurteilung und Bestätigung der Kategorisierung dar.

In der vorliegenden Arbeit analysierte ich ausschließlich Bilddaten desselben Scanners (Siemens Biograph 2 PET/CT, Siemens Healthineers, Erlangen, Germany). Abschließen ist anzumerken, dass die Einbeziehung unterschiedlicher PET/CT-Geräte und Abläufe in die Analyse sowie deren Einfluss auf die Ergebnisse der Vorhersage in weiteren Studien untersucht werden sollte.

1.5 Zusammenfassung

Die initiale Fragestellung dieser Arbeit war, inwieweit ein Algorithmus in der Lage ist, *hotspots* in der ⁶⁸Ga-PSMA-PET/CT bei Patienten mit fortgeschrittenem Prostatakarzinom als pathologisch und somit maligne zu erkennen.

Ich markierte, analysierte und kategorisierte 2452 Läsionen von 72 Patienten. So entstanden die Hauptkategorien "pathologisch" und "physiologisch". Letztere umfasst mehrere Untergruppen, die sich auf die verschiedenen Organe beziehen, in denen eine vermehrte Anreicherung von ⁶⁸Ga-PSMA als benigne zu bewerten ist.

Es folgte der Vergleich von drei ML-Algorithmen durch *cross-validation*. Der *Extra Trees classifier* stellte sich im Vergleich zu *random forest* und *SVM* als präziser heraus.

Der nächste Schritt war das Training des Algorithmus mithilfe der oben beschriebenen *hotspots* aus der Trainingskohorte. *Extra Trees* erreichte nach der Trainingsphase mit 72 Patienten eine AUC von 0,98, eine Sensitivität von 0,97 und eine Spezifität von 0,83.

Anschließend quantifizierten wir unsere Ergebnisse in Hinblick auf unsere Fragestellung. Dazu unterzogen wir den Algorithmus einem Test, bestehend aus 331 *hotspots* bei 15 Patienten. Diese Patienten ähnelten der Trainingskohorte in Alter, Gleason Score und PSA-Wert.

Der Einsatz von ML im Sinne eines überwachten Lernens auf ⁶⁸Ga-PSMA-PET/CT Scans erreichte überzeugende Ergebnisse in der Identifizierung von malignen Läsionen bei Prostatakarzinom-Patienten. Dies unterstreicht die Bedeutung der computerunterstützten Diagnose und Therapiefindung, die sich als zukunftsweisende Richtung in der Medizin erweist. Bei jeder Anwendung von ML sind jedoch auch Einschränkungen im klinischen Kontext zu berücksichtigen.

Verbesserungen sind im Bereich der Klassifizierung kleiner Speicheldrüsen und in der Validierung der Aufnahme des Markierungsstoffes in der Prostata selbst anzustreben. Um andere Einschränkungen und Einflüsse wie Variabilität in der Bildanalyse und bestimmte Eigenschaften der PET-Scanner auszuschließen, sollten sich weiterführende, möglichst multizentrische Studien anschließen.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Alonso-Betanzos A, Bolón-Canedo V. Big-Data Analysis, Cluster Analysis, and Machine-Learning Approaches. Advances in experimental medicine and biology 2018; 1065: 607– 626

Alpaydin E. Maschinelles Lernen. München, Wien: De Gruyter Oldenbourg 2019

Anand Shroff. 2017. For machine learning to be adopted in healthcare, know its limitations. Verfügbar unter https://medcitynews.com/2017/10/machine-learning-adopted-healthcare-know-risks/ (Zugriffsdatum: 27.09.2021)

AWMF, Deutsche Krebsgesellschaft, Deutsche Krebshilfe. 2021. S3-Leitlinie Prostatakarzinom. Verfügbar unter https://www.leitlinienprogrammonkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Prostatatkarzinom/Version_6/L L_Prostatakarzinom_Langversion_6.0.pdf (Zugriffsdatum: 15.09.2021)

Bluemel C, Krebs M, Polat B, Linke F, Eiber M, Samnick S, Lapa C, Lassmann M, Riedmiller H, Czernin J, Rubello D, Bley T, Kropf S, Wester H-J, Buck AK, Herrmann K. 68Ga-PSMA-PET/CT in Patients With Biochemical Prostate Cancer Recurrence and Negative 18F-Choline-PET/CT. Clinical nuclear medicine 2016; 41: 515–521

Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. Current Applications and Future Impact of Machine Learning in Radiology. Radiology 2018; 288: 318–328

Deo RC. Machine Learning in Medicine. Circulation 2015; 132: 1920–1930

Erle A, Moazemi S, Lütje S, Essler M, Schultz T, Bundschuh RA. Evaluating a Machine Learning Tool for the Classification of Pathological Uptake in Whole-Body PSMA-PET-CT Scans. Tomography (Ann Arbor, Mich.) 2021; 7: 301–312 García Garzón JR, Arcocha Torres M de, Delgado-Bolton R, Ceci F, Alvarez Ruiz S, Orcajo Rincón J, Caresia Aróztegui AP, García Velloso MJ, García Vicente AM. La PET/TC con 68Ga-PSMA en el cáncer de próstata. Revista espanola de medicina nuclear e imagen molecular 2018; 37: 130–138

Giesel FL, Knorr K, Spohn F, Will L, Maurer T, Flechsig P, Neels O, Schiller K, Amaral H, Weber WA, Haberkorn U, Schwaiger M, Kratochwil C, Choyke P, Kramer V, Kopka K, Eiber M. Detection Efficacy of 18F-PSMA-1007 PET/CT in 251 Patients with Biochemical Recurrence of Prostate Cancer After Radical Prostatectomy. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 2019; 60: 362–368

Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. Journal of internal medicine 2018; 284: 603–619

Heck MM, Retz M, Tauber R, Knorr K, Kratochwil C, Eiber M. Radionuklidtherapie des Prostatakarzinoms mittels PSMA-Lutetium. Der Urologe. Ausg. A 2017; 56: 32–39

Hummel H-D, Kufer P, Grüllich C, Deschler-Baier B, Chatterjee M, Goebeler M-E, Miller K, Santis M de, Loidl WC, Buck A, Wittemer-Rump S, Koca G, Boix O, Doecke W-D, Stienen S, Sayehli C, Bargou RC. Phase 1 study of pasotuxizumab (BAY 2010112), a PSMA-targeting Bispecific T cell Engager (BiTE) immunotherapy for metastatic castration-resistant prostate cancer (mCRPC). JCO 2019; 37: 5034

Hummel H-D, Kufer P, Grüllich C, Seggewiss-Bernhardt R, Deschler-Baier B, Chatterjee M, Goebeler M-E, Miller K, Santis M de, Loidl W, Dittrich C, Buck A, Lapa C, Thurner A, Wittemer-Rump S, Koca G, Boix O, Döcke W-D, Finnern R, Kusi H, Ajavon-Hartmann A, Stienen S, Sayehli CM, Polat B, Bargou RC. Pasotuxizumab, a BiTE® immune therapy for castration-resistant prostate cancer: Phase I, dose-escalation study findings. Immunotherapy 2021; 13: 125–141

Jones W, Griffiths K, Barata PC, Paller CJ. PSMA Theranostics: Review of the Current Status of PSMA-Targeted Imaging and Radioligand Therapy. Cancers 2020; 12

Basiswissen Radiologie. Kahl-Scholz M, Vockelmann C, Reg. Berlin: Springer 2017 Kamat NV, Yu EY, Lee JK. BiTE-ing into Prostate Cancer with Bispecific T-cell Engagers. Clinical cancer research : an official journal of the American Association for Cancer Research 2021; 27: 2675–2677

Khreish F, Ghazal Z, Marlowe RJ, Rosar F, Sabet A, Maus S, Linxweiler J, Bartholomä M, Ezziddin S. 177 Lu-PSMA-617 radioligand therapy of metastatic castration-resistant prostate cancer: Initial 254-patient results from a prospective registry (REALITY Study). European journal of nuclear medicine and molecular imaging 2021

Kratochwil C, Bruchertseifer F, Giesel FL, Weis M, Verburg FA, Mottaghy F, Kopka K, Apostolidis C, Haberkorn U, Morgenstern A. 225Ac-PSMA-617 for PSMA-Targeted α-

Radiation Therapy of Metastatic Castration-Resistant Prostate Cancer. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 2016; 57: 1941–1944

Kuten J, Fahoum I, Savin Z, Shamni O, Gitstein G, Hershkovitz D, Mabjeesh NJ, Yossepowitch O, Mishani E, Even-Sapir E. Head-to-Head Comparison of 68Ga-PSMA-11 with 18F-PSMA-1007 PET/CT in Staging Prostate Cancer Using Histopathology and Immunohistochemical Analysis as a Reference Standard. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 2020; 61: 527–532

Lin C-Y, Lee M-T, Lin C-L, Kao C-H. Comparing the Staging/Restaging Performance of 68Ga-Labeled Prostate-Specific Membrane Antigen and 18F-Choline PET/CT in Prostate Cancer: A Systematic Review and Meta-analysis. Clinical nuclear medicine 2019; 44: 365–376

Moazemi S, Erle A, Khurshid Z, Lütje S, Muders M, Essler M, Schultz T, Bundschuh RA. Decision-support for treatment with 177Lu-PSMA: machine learning predicts response with high accuracy based on PSMA-PET/CT and clinical parameters. Annals of translational medicine 2021a; 9: 818

Moazemi S, Erle A, Lütje S, Gaertner FC, Essler M, Bundschuh RA. Estimating the Potential of Radiomics Features and Radiomics Signature from Pretherapeutic PSMA-PET-CT Scans and Clinical Data for Prediction of Overall Survival When Treated with 177Lu-PSMA. Diagnostics (Basel, Switzerland) 2021b; 11

Müller AC, Guido S. Einführung in Machine Learning mit Python. Heidelberg: O'Reilly 2017

National Cancer Institute. 2021. Cancer Stat Facts: Prostate Cancer (Zugriffsdatum: 29.09.2021)

National Library of Medicine. 2021. pubmed.gov (Zugriffsdatum: 21.09.2021)

Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. The New England journal of medicine 2016; 375: 1216–1219

Petersen LJ, Nielsen JB, Langkilde NC, Petersen A, Afshar-Oromieh A, Souza NM de, Paepe K de, Fisker RV, Arp DT, Carl J, Haberkorn U, Zacho HD. 68Ga-PSMA PET/CT compared with MRI/CT and diffusion-weighted MRI for primary lymph node staging prior to definitive radiotherapy in prostate cancer: a prospective diagnostic test accuracy study. World journal of urology 2020; 38: 939–948

Rahbar K, Afshar-Oromieh A, Jadvar H, Ahmadzadehfar H. PSMA Theranostics: Current Status and Future Directions. Molecular imaging 2018; 17: 1536012118776068

Rahbar K, Ahmadzadehfar H, Kratochwil C, Haberkorn U, Schäfers M, Essler M, Baum RP, Kulkarni HR, Schmidt M, Drzezga A, Bartenstein P, Pfestroff A, Luster M, Lützen U, Marx M, Prasad V, Brenner W, Heinzel A, Mottaghy FM, Ruf J, Meyer PT, Heuschkel M, Eveslage M, Bögemann M, Fendler WP, Krause BJ. German Multicenter Study Investigating 177Lu-PSMA-617 Radioligand Therapy in Advanced Prostate Cancer Patients. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 2017; 58: 85–90

Reig B, Heacock L, Geras KJ, Moy L. Machine learning in breast MRI. Journal of magnetic resonance imaging : JMRI 2020; 52: 998–1018

Sachpekidis C, Bäumer P, Kopka K, Hadaschik BA, Hohenfellner M, Kopp-Schneider A, Haberkorn U, Dimitrakopoulou-Strauss A. 68Ga-PSMA PET/CT in the evaluation of bone metastases in prostate cancer. European journal of nuclear medicine and molecular imaging 2018; 45: 904–912

Silver DA, Pellicer I, Fair WR, Heston WD, Cordon-Cardo C. Prostate-specific membrane antigen expression in normal and malignant human tissues. Clinical cancer research : an official journal of the American Association for Cancer Research 1997; 3: 81–85

Tan N, Bavadian N, Calais J, Oyoyo U, Kim J, Turkbey IB, Mena E, Davenport MS. Imaging of Prostate Specific Membrane Antigen Targeted Radiotracers for the Detection of Prostate Cancer Biochemical Recurrence after Definitive Therapy: A Systematic Review and Meta-Analysis. The Journal of urology 2019; 202: 231–240

WHO-International Agency for Research on Cancer. 2021. Global Cancer Observatory. Verfügbar unter https://gco.iarc.fr/ (Zugriffsdatum: 29.09.2021)

Wondergem M, Jansen BHE, van der Zant FM, van der Sluis TM, Knol RJJ, van Kalmthout LWM, Hoekstra OS, van Moorselaar RJA, Oprea-Lager DE, Vis AN. Early lesion detection with 18F-DCFPyL PET/CT in 248 patients with biochemically recurrent prostate cancer. European journal of nuclear medicine and molecular imaging 2019; 46: 1911–1918

Zhou J, Gou Z, Wu R, Yuan Y, Yu G, Zhao Y. Comparison of PSMA-PET/CT, choline-PET/CT, NaF-PET/CT, MRI, and bone scintigraphy in the diagnosis of bone metastases in patients with prostate cancer: a systematic review and meta-analysis. Skeletal radiology 2019; 48: 1915–1924

2. Veröffentlichung





Article Evaluating a Machine Learning Tool for the Classification of Pathological Uptake in Whole-Body PSMA-PET-CT Scans

Annette Erle ^{1,†}, Sobhan Moazemi ^{1,2,*,†}, Susanne Lütje ¹, Markus Essler ¹, Thomas Schultz ^{2,3}, and Ralph A. Bundschuh ¹

- ¹ Department of Nuclear Medicine, University Hospital Bonn, 53127 Bonn, Germany; annette.erle@t-online.de (A.E.); susanne.luetje@ukbonn.de (S.L.); markus.essler@ukbonn.de (M.E.); ralph.bundschuh@ukbonn.de (R.A.B.)
- ² Department of Computer Science, University of Bonn, 53115 Bonn, Germany; schultz@cs.uni-bonn.de
- ³ Bonn-Aachen International Center for Information Technology (B-IT), 53115 Bonn, Germany
- * Correspondence: s.moazemi@ukbonn.de; Tel.: +49-2282-8715-538
- + Both authors contributed in equal to the manuscript.

Abstract: The importance of machine learning (ML) in the clinical environment increases constantly. Differentiation of pathological from physiological tracer-uptake in positron emission to mography/computed tomography (PET/CT) images is considered time-consuming and attention intensive, hence crucial for diagnosis and treatment planning. This study aimed at comparing and validating supervised ML algorithms to classify pathological uptake in prostate cancer (PC) patients based on prostate-specific membrane antigen (PSMA)-PET/CT. Retrospective analysis of ⁶⁸Ga-PSMA-PET/CTs of 72 PC patients resulted in a total of 77 radiomics features from 2452 manually delineated hotspots for training and labeled pathological (1629) or physiological (823) as ground truth (GT). As the held-out test dataset, 331 hotspots (path.:128, phys.: 203) were delineated in 15 other patients. Three ML classifiers were trained and ranked to assess classification performance. As a result, a high overall average performance (area under the curve (AUC) of 0.98) was achieved, especially to detect pathological uptake (0.97 mean sensitivity). However, there is still room for improvement to detect physiological uptake (0.82 mean specificity), especially for glands. The ML algorithm applied to manually delineated lesions predicts hotspot labels with high accuracy on unseen data and may be an important tool to assist in clinical diagnosis.

Keywords: prostate cancer (PC); prostate specific membrane antigen (PSMA); positron emission tomography (PET); computed tomography (CT); radiomics features (RFs); machine learning (ML)

1. Introduction

Prostate Cancer is one of the leading malignancies in the whole world, with 1.3 million new cases in 2018. It is the second most common cancer in men worldwide [1] and is the fifth leading cause of cancer death [2]. The number of deaths is estimated to rise by 105.6% by 2040 [2]. While the overall 5-year-survival-rate reaches 98.0%, it decreases to 30.5% once metastases occur [3].

In order to diagnose the disease, detect possible progression and monitor therapy response, body scans are indispensable. Hybrid positron emission tomography/computed tomography (PET/CT) scans depict anatomical data on the one hand and functional information of tissues on the other hand. As a favorable biomarker for PET in prostate cancer patients, radiolabeled ligands to prostate-specific membrane antigen (PSMA) are used, being considered in many studies [4]. PSMA is expressed in prostate cells in very low concentration and overexpressed in an increasing degree on prostate cancer cells [5]. Therefore, it allows distinguishing between benign and malignant tissues. For PET diagnosis, small molecules targeting PSMA and labeled with the positron emitters, gallium-68 (⁶⁸Ga) [6] or



Citation: Erle, A.; Moazemi, S.; Lütje, S.; Essler, M.; Schultz, T.; Bundschuh, R.A. Evaluating a Machine Learning Tool for the Classification of Pathological Uptake in Whole-Body PSMA-PET-CT Scans. *Tomography* 2021, 7, 301–312. https://doi.org/ 10.3390/tomography7030027

Academic Editor: Benjamin M. Ellingson

Received: 28 May 2021 Accepted: 27 July 2021 Published: 29 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). fluorine-18 (¹⁸F) [7], are utilized. The combination of PSMA-PET and CT allows allocating metastasis accurately [8,9].

So far, analysis of PSMA-PET/CT image data, performed by specialists in nuclear medicine, is mainly done manually based on experience. This has several disadvantages, being time-consuming and error-prone with high numbers of inter- and intra-observer variability [10]. Here, the use of computer-aided diagnosis (CAD) has several advantages: it can improve diagnostic precision, accelerates analysis, facilitates the clinical workflow, lowers human resource costs, and even may predict prognosis, and last but not least, if trained with proper datasets, can compensate for inter- and intra-observer variability [11]. Machine learning (ML) algorithms base their analysis on very large amounts of unstructured information ("big data" [12]), which allows the recognition of complex patterns. Especially in imaging, their application has shown to be very successful [13–15]. In hybrid imaging, the implementation of artificial intelligence (AI) is a promising field with application in a wide range of clinical sectors, e.g., Morbus Alzheimer in neurology, lung cancer, multiple myeloma, and prostate cancer in oncology [10,11]. Analysis of textural parameters in standard clinical image data allows performing segmentation and characterization of tissues as well as being useful in the fields of prediction and prognosis, leading the way to individually tailored therapies in the future [16–18]. In the field of prostate cancer, ML methods are furthermore already used and growing, e.g., in treatment, histopathology and genetics [13,14]. Khurshid et al. have shown that the two textural heterogeneity parameters, entropy and homogeneity, correlate with pre/post-therapy prostate-specific antigen (PSA) levels. A higher level of heterogeneity seems to predict a better response to PSMA therapy and may in the future allow a pre-therapeutic selection of responders to the treatment [19]. Moazemi et al. have consecutively shown that these parameters also have prognostic potential for the overall survival of prostate cancer patients [20]. In summary, the software could take over and improve the whole process in the future, enabling specialists to focus on more important tasks. For such CAD, including decision support algorithms, at first, the detection of pathological lesions is necessary, followed by an analysis of the radiomics features in these lesions. Although the automatic segmentation of high tracer uptake would be an essential step towards such a clinical decision support tool, in this study, we focus on the automated classification of pathological vs physiological uptake using radiomics features from manually segmented hotspots and leave the automated segmentation as future work. When present, such an automated system would enhance the procedure of the management of PC patients in terms of time and effort.

Therefore, in the present work, we further compare and evaluate supervised ML-based algorithms for classifying hotspots in PSMA PET/CT, which have shown their potential before [21]. Additionally, to verify the significance of the existing algorithm, the training cohort was gradually extended to find out with how many subjects our model would generalize. The aim of this study was to quantify the accuracy of the algorithm as applied to unseen sets of data, especially focusing on enhancing the true classification of hotspots with physiological uptake. To this end, first, the patients' scans are manually annotated to provide data. Then, python software undertakes the task of classifying uptakes into two categories (pathological vs physiological). Finally, the output is verified and reviewed with an independent test cohort.

2. Materials and Methods

2.1. Patients and Volume of Interest (VoI) Delineation

Data of 87 patients with histologically proven prostate cancer were included in this analysis. All patients received a PSMA-PET/CT examination due to clinical reasons, either for staging or treatment control. PET/CT examinations were performed 40 to 80 min after the intravenous injection from 98 to 159 MBq in-house produced ⁶⁸Ga-HBED-CC PSMA using a Siemens Biograph 2 PET/CT machine (Siemens Healthineers, Erlangen, Germany). First, a low-dose CT (16 mAS, 130 kV) was performed from the skull to mid-thigh, followed by the PET imaging over the same area with 3 or 4 min per bed position

depending on the weight of the patient. PET data were reconstructed in 128×128 matrices with 5 mm slice thickness, while CT data were reconstructed in 512×512 matrices with 5 mm slice thickness. For PET image reconstruction, the attenuation-weighted ordered subsets expectation-maximization algorithm implemented by the manufacturer was used, including attenuation and scatter correction based on the CT data. Additionally, applied by the manufacturer, a 5 mm post-reconstruction Gaussian filtering was used for smoothing of all the input images prior to the ML analyses. All patients gave written and informed consent to the diagnostic procedure. Due to the retrospective character of the data analysis, an ethical statement was waived by the institutional ethical review board according to the professional regulations of the medical board of Nordrhein-Westfalen, Germany.

A total of 72 patients were assigned as the training dataset. The patients' average age was 71 (range: 48–87), and the average Gleason score was 8 (range: 6–10) (Table 1). The PET/CT images data were analyzed using Interview Fusion Software by Mediso Medical Imaging (Budapest, Hungary) [22]. All the hotspots have been identified based on fused PET and CT data. Volumes of interest (VoIs) were manually delineated with a brush tool in the PET images slice by slice (Figure 1). The criteria to choose an uptake was the visible tracer uptake without any predefined threshold. In a second step, the hotspots were classified as pathological or physiological, corresponding to the location they were situated in. The hotspots included primary prostate cancer and metastases in the skeletal system, lymph nodes, as well as physiological uptake in kidneys, liver, glands, gastrointestinal tract (gut) etc. A total number of 2452 hotspots were marked and then categorized as either pathological (total of 1629) or physiological (total of 823).

Table 1. The age, Gleason score, and PSA level ranges for the patients in the training and test cohorts.

Parameter	aining Grou	ng Group Test Cohort				
Name	Average	Min	Max	Average	Min	Max
Age	71	48	87	77	63	87
Gleason Score	8	6	10	8	6	10
PSA (ng/mL)	438.72	4.73	5910.0	660.62	1.2	5400.0



Figure 1. A sample screenshot from the hotspots' delineation step using InterView Fusion Software (Mediso, Budapest, Hungary) [22]. The hotspots are identified and delineated slice by slice as 3D volumes of interests (VoIs) by an experienced nuclear medicine physician (the hotspot names translation from German: Niere = kidney, Darm = gut. Moreover, Met stands for metastasis).

In total, 15 remaining patients with similar ranges of age, Gleason score, and PSA level as the patients in the training cohort were assigned to the test group for testing the ML algorithm (Table 1). First, the scans were again analyzed in the same way with Interview Fusion software as described above. For every patient, 5 to 10 pathological hotspots were delineated and tracer uptake in all glands; 5 physiological hotspots and uptake in the liver. Beforehand, glands proved to be difficult for the algorithm to classify whether physiological or pathological. This analysis resulted in 331 hotspots with 128 pathological and 203 physiological lesions.

For each hotspot, a total of 77 radiomics features were calculated using InterView FUSION software (Mediso Medical Imaging, Budapest, Hungary). The features include first/higher order statistics, textural heterogeneity parameters, and zone/run length statistics features. The complete list of the features is provided in Table 2.

Table 2. A list of the radiomics features from both positron emission tomography (PET) and computed tomography (CT) modalities. Please note that the metabolic tumor volume (MTV) is PET-specific. This table is already published in [20].

First or Higher Order Statistics	Shape and Size	Textural	Volumetric Zone Length Statistics	Volumetric Run Length Statistics
Deviation	Max. Diameter	Entropy	Short Zone Emphasis	Short Run Emphasis
Mean		Homogeneity	Long Zone Emphasis	Long Run Emphasis
Max		Correlation	Low Grey-Level Zone	Low Grey-Level Run
IVIAX		Correlation	Emphasis	Emphasis
Min		Contract	High Grey-Level Zone	High Grey-Level Run
IVIIII		Contrast	Emphasis	Emphasis
Sum		Size Variation	Short Zone Low	Short Run Low Grey-Level
Sum		Size variation	Grey-Level Emphasis	Emphasis
DET MTV		Intensity Variation	Short Zone High	Short Run High Grey-Level
		intensity variation	Grey-Level Emphasis	Emphasis
Vurtosis		Contropos	Long Zone Low	Long Run Low Grey-Level
Kurtosis		Coarseness	Grey-Level Emphasis	Emphasis
		Buckmoord	Long Zone High	Long Run High Grey-Level
		Dusyness	Grey-Level Emphasis	Emphasis
				Grey-Level
		Complexity	Zone Percentage	Non-Uniformity Run
		-	-	Length Non-Uniformity
				Run Percentage

2.2. Training and Classification

For the ML analyses, the question of pathological versus physiological uptake was mapped to the so-called supervised ML problem [23]. This sort of ML algorithm is applied when part of the study cohort already includes complete information on input variables (in our case, PET/CT hotspots and their corresponding radiomics features) as well as ground truth labels (in our case, pathological vs physiological). Thus, for training and classification purposes, an in-house developed software in Python V.3.5 was used. Initially, we used a 30-subjects subset of the training dataset of 72 subjects to pre-set, tune, and compare our machine learning classifiers from SciKitLearn library [24] (linear kernel support vector machine (SVM), as well as ExtraTrees [25] and random forest as classifiers based on decision trees [26]), which already showed their significance [21].

In the first round of training, cross-validation (CV) is applied with KFold with 3 folds to tune hyperparameters of the classifiers and identify the best performing one. To this end, at each CV step, the C and Gamma parameters of the linear SVM as well as the min_sample_leaf and max_depth of the decision tree-based algorithms are tuned using grid search. For the hyperparameter tuning, standard ranges of the hyperparameters are applied. For example, for C, the range from 2^{-5} to 2^{13} is used, and for max_depth, the range from 1 to 10 is used for the grid search. This has resulted in the best combination of the hyperparameters for each classifier and helped to identify the best classifier based on

the performance metrics AUC, sensitivity (SE), and specificity (SP). We further quantified the performances of the classifiers on the test cohort to end up with the best algorithm, which was Extra Trees. From this point, we used Extra Trees with its tuned hyperparameters (n_estimators = 250, max_depth = 20, min_samples_leaf = 1) to investigate how the algorithm would generalize as the size of the training cohort increases.

For the next step to assess the generalizability of the algorithms, we started with the first initial cohort of 30 patients for training the algorithm. Then, we added the data from the second training cohort (42 patients), one patient at a time and with a randomized order (thus, the sizes of the training subsets varied from 30 to 72). It means, in each training step, a random subset of the combination of the two training cohorts was chosen, and each time the size of the subset was increased by one patient. Furthermore, we repeated the classification task 100 times with a bootstrapping approach to calculate the accuracy measures at each training step. As we aimed at assessing the performance of our algorithm on unseen data, in each step, we calculated the prediction accuracies (AUC, sensitivity (SE), and specificity (SP)) on the validation set (the hold-out set with 15 subjects). Finally, we report the mean and standard deviation (std) of the accuracy metrics to give an overview of how increasing the size of the training cohort affects the classification performance metrics when trained by the training cohorts of 30 to 72 subjects and tested by the test cohort of 15 subjects. The performance metrics were calculated at each training step based on the prediction scores of the ML classifier as trained by the training cohort and tested by the test cohort, then averaged along 100 bootstraps.

To minimize the risk of overfitting, first, the dataset including feature vectors of training and test subjects were normalized, using the MinMax standardization method. This method maps the input variables into the range between 0 and 1 to compensate for inconsistent variable ranges. As a result, variables with very large or very small values would not affect the classifier performance. For this study, we chose to apply both cross-validations to identify the best classifier and bootstrapping with replacement and resampling on the training set to better estimate the population statistics.

3. Results

3.1. Pathological and Physiological Hotspots

In the 72 patients' scans of the training cohort, 2452 hotspots have been identified and marked. Out of these, 823 hotspots have been classified as physiological versus 1634 as pathological. In the 15 patients of the control cohort, 331 areas with increased uptake have been defined. Table 3 shows the detailed distribution of hotspots in the training cohort, and Table 4 shows the summarized distribution of the hotspots in the train and test cohorts. In the training cohort of 72 patients, the category "others" denotes 33 lesions in other organs such as the spleen and lung.

Table 3. The distribution of hotspots throughout the training cohort subjects.

Patient ID	Metastasis	Bladder	Gut	Kidney	Ureter	Parotid Gland	Lacrimal Gland	Sub-Mandibular Gland	Sub-Lingual Gland	Prostate	Liver	Other
$ \begin{array}{c} 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\\13\\14\\15\\16\\17\\18\\19\\20\\21\\22\\23\\24\\25\\26\\27\\28\\29\\30\\31\\32\\33\\34\\35\\36\\37\\38\\39\\40\\41\end{array} $	$\begin{array}{c} 11\\ 26\\ 21\\ 25\\ 15\\ 28\\ 30\\ 20\\ 16\\ 25\\ 11\\ 19\\ 25\\ 34\\ 11\\ 35\\ 18\\ 17\\ 7\\ 27\\ 14\\ 22\\ 18\\ 19\\ 31\\ 12\\ 29\\ 43\\ 13\\ 29\\ 13\\ 24\\ 24\\ 22\\ 13\\ 30\\ 30\\ 15\\ 21\\ 25\\ 18\end{array}$	1 1 0 0 1 0 1 1 0 0 1 1 1 0 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1	$ \begin{array}{c} 1\\1\\1\\0\\2\\1\\1\\1\\1\\1\\2\\0\\1\\1\\1\\1\\1\\1\\0\\2\\1\\1\\1\\0\\2\\1\\1\\1\\1$	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	$ \begin{array}{c} 2\\ 0\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 1\\ 2\\ 0\\ 0\\ 1\\ 2\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$	$ \begin{array}{c} 2\\ 0\\ 0\\ 2\\ 1\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 0\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\ 2\\$	$ \begin{array}{c} 2 \\ 1 \\ 2 \\ 0 \\ 0 \\ 2 \\ 2 \\ 1 \\ 1 \\ 2 \\ 0 \\ 2 \\ 2 \\ 0 \\ 0 \\ 2 \\ $	O 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 0 2 0 0 0 0 0 0 0 2 0 <td< td=""><td>5 0 2 0 1 1 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 2 1 0 0 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 1 2 0 0 2 2 2 1 0 0 0 1 2 0 0 1 2 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 1 2 0 0 0 0 2 2 2 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>$\begin{array}{c} 1\\ 1\\ 0\\ 0\\ 0\\ 1\\ 0\\ 1\\ 0\\ 1\\ 0\\ 1\\ 1\\ 0\\ 0\\ 1\\ 0\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$</td><td>1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>$\begin{array}{c} 2\\ 0\\ 0\\ 0\\ 4\\ 0\\ 0\\ 0\\ 3\\ 2\\ 0\\ 1\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$</td></td<>	5 0 2 0 1 1 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 2 1 0 0 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 2 2 1 0 0 1 2 0 0 2 2 2 1 0 0 0 1 2 0 0 1 2 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 2 2 2 1 0 0 0 0 1 2 0 0 0 0 2 2 2 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	$ \begin{array}{c} 1\\ 1\\ 0\\ 0\\ 0\\ 1\\ 0\\ 1\\ 0\\ 1\\ 0\\ 1\\ 1\\ 0\\ 0\\ 1\\ 0\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$	1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	$\begin{array}{c} 2\\ 0\\ 0\\ 0\\ 4\\ 0\\ 0\\ 0\\ 3\\ 2\\ 0\\ 1\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$
41 42 43 44 45 46 47	20 23 36 11 34 19	1 1 1 1 1 1	0 1 1 1 1 1	2 2 2 2 2 2 2 2	0 0 0 0 2 0	2 2 2 2 2 2 2 2	2 2 2 2 2 2 2 2	2 2 2 2 2 2 2 2	0 0 1 1 0 0	0 0 0 0 0 0	0 1 1 0 0 1	0 0 0 2 0 0
48 49 50	15 7 28	1 1 1	1 1 1	2 2 2	2 2 2	2 2 2	2 2 2	2 2 2	2 2 0	0 0 0	0 0 1	0 1 0

Table 3. Cont.

Patient ID	Metastasis	Bladder	Gut	Kidney	Ureter	Parotid Gland	Lacrimal Gland	Sub-Mandibular Gland	Sub-Lingual Gland	Prostate	Liver	Other
51	10	1	1	2	0	2	2	2	2	0	0	1
52	26	1	1	1	0	2	2	2	2	0	0	0
53	14	1	1	2	0	2	2	2	0	0	1	0
54	23	1	1	2	2	2	2	2	2	0	0	2
55	31	1	1	2	3	2	2	2	2	0	0	0
56	12	1	1	2	0	2	2	2	2	0	0	0
57	27	1	1	2	0	2	2	2	2	0	0	0
58	45	1	1	2	0	2	2	2	1	0	1	0
59	10	1	1	2	0	2	2	2	2	0	0	0
60	26	1	1	2	0	4	2	2	2	0	0	0
61	20	1	1	2	1	2	2	2	2	0	1	1
62	29	1	1	2	0	2	1	2	0	0	0	0
63	29	1	1	2	0	2	2	2	2	0	0	0
64	26	1	3	2	0	2	2	2	2	0	1	0
65	20	1	1	2	0	2	2	2	2	0	1	0
66	19	0	1	2	0	2	2	2	1	0	1	0
67	47	1	1	2	1	2	2	2	0	0	0	0
68	16	1	2	2	0	2	0	2	1	0	0	0
69	25	1	2	0	0	2	2	2	2	0	0	0
70	21	0	1	2	0	2	2	2	2	0	0	0
71	42	1	1	2	0	2	2	2	0	0	1	0
72	31	1	1	2	0	2	2	2	2	0	0	0
total	1620	58	76	140	57	122	115	107	71	14	39	33

Table 4. The distribution of the hotspots over different body organs in the training and test cohorts. (* the others category refers to the rest of organs in which hotspots were located).

Hotspot Category	Training Cohort 72 Patients	Test Cohort 15 Patients
Metastases	1620	128
Bladder	58	13
Kidney	140	30
Salivary Gland	415	111
Prostate	14	-
gut	76	33
liver	39	15
ureter	57	1
Others *	33	-
Total of hotspots	2452	331

3.2. Patient Validation

As a result of the cross-validation step, the Extra Trees classifier outperformed other classifiers (see Figure 2). The Extra Trees classifier trained with the data of 30 patients resulted in an AUC of 0.95, a sensitivity of 0.95, and a specificity of 0.80. As expected, extending the training cohort by incrementing the sample size, one patient after another, resulted in improved accuracy measures until it reached its overall maximum as the data from the 72nd subject was imported (0.98 AUC, 0.97 sensitivity, and 0.83 specificity). Figure 2 gives an overview of how different classifiers performed on the first training cohort, and Table 5 shows the mean and standard deviation of the accuracy measures

Tomography 2021, 7,

along all the training steps. Based on the training data of 72 subjects and the validation cohort of 15 subjects, 125 of the 128 lesions defined by the reader as pathological were identified as pathological as well, which means a sensitivity of 0.97. Additionally, the physiological uptake was classified with high precision. While the algorithm showed accurate predictions of liver, kidneys, GUT, etc., specificity regarding salivary glands showed to be more difficult with a figure of 0.82. Especially in sublingual and lacrimal glands, there was a high rate of false positives (9/19 and 7/19). Other glands were easier to identify as physiological, resulting in a total of only 2 (submandibular gland) and 1 (parotid gland) false positives. The complete overview is presented in Table 6. As pathological prostate uptake was only present in 14 subjects from the training cohort, we could not analyze the prediction performance on the test cohort for this category. However, analyzing the 14 prostate hotspots from the training cohort, we achieved a sensitivity of 0.92 (13/14⁹) true positives).

34





TTold 5.5T Klomenon and statandard downiatic (st(1) 4) halves the hear an average of the there is the second state of the seco

Accuracy Metr	curacy Metric	Mean	Mean	Std	Std
AUC	AUC	0.98	0.98	0.002	0.002
Sensitivity	Sensitivity	0.97	0.97	0.004	0.004
Specificity	Specificity	0.82	0.82	0.02	0.02

Table 6. The results of the predictions on the test cohort. (GUT: gastrointestinal tract).

Category	1/Pathological	0/Physiological	Total	Specificity
bladder	0	13	13	1.00
glands	19	92	111	0.82
GUT	1	32	33	0.97
liver	0	15	15	1.00

Category	1/Pathological	0/Physiological	Total	Specificity
bladder	0	13	13	1.00
glands	19	92	111	0.82
GUT	1	32	33	0.97
liver	0	15	15	1.00
kidney	0	30	30	1.00
ureter	0	1	1	1.00
metastases	125	3	128	0.97

Table 6. The results of the predictions on the test cohort. (GUT: gastrointestinal tract).

4. Discussion

The application of ML-based image analysis would implicate several benefits for patients. First of all, this examination is fast and non-invasive. Second of all, it could be used for therapy prediction and differentiation between low and high-risk patients [27] and for prediction of overall survival [20], as already published before. It could furthermore be the base of individual therapy and treatment tailoring in the future, combining the technique with analysis of textural parameters. Furthermore, the presented ML-based approach and the outcoming results confirm the potential of the ML methods for the analysis of newly day-to-day arriving patients' data.

The problem with the smaller set of data was that more physiological hotspots were recognized as pathological, i.e., a high rate of false positives. Results with the bigger set of data have shown to be more specific though. However, the obtained prediction accuracies after increasing the training cohort size showed the robustness of the algorithms. In addition, it is important to note that we have a very high number of hotspots per patient (34 hotspots per patient on average). Therefore, even with a low number of patients we have, significant numbers of lesions.

As a matter of fact, overfitting is an important concern in ML studies, especially when the number of features exceeds the number of samples. In this study, we included 2452 hotspots data (n_samples = 2452) and 77 radiomics features (n_features = 77). We also applied MinMax standardization. We had already applied cross-validation in the previous study [21] to identify the best performing ML classifier and tune its hyperparameters. In this study, we applied Bootstrapping to better estimate the population statistics. To conclude, all these precautions had been made to avoid overfitting.

In this analysis, we could also identify in which locations the most problems appear in the AI algorithm. In fact, the categorization of glands proved to be difficult with a specificity of just 0.82. 19 out of 111 glands were identified as metastasis, especially sublingual (9/19) and lacrimal (7/19) glands. The reason why the ML algorithm performs poorer on glands might be caused by the fact that, from the data-driven point of view, the radiomics features of the glands seem to be in a similar range as for the pathological uptake. Therefore, this must be a topic of further improvement of the algorithm. However, this limitation of the algorithm seems to be acceptable for this feasibility study as the head is only in very far spread tumor disease, a typical location of metastases. For other organ uptakes as kidney and bladder, a specificity of 1.0 was reached. Even uptake in the GUT approached a result just as high with a specificity of 0.97. The performance of the ML classifier to classify pathological uptake was also quantified as high (0.92 sensitivity). However, as no pathological uptake was annotated in the test cohort, we could not analyze it on the test step.

As the goal of the presented study was to show the feasibility and power of the lesion classification in PSMA-PET data, so far, we used manual segmentation. For fast data processing in clinical routine in future segmentation needs to be performed in an automatic or at least semi-automatic manner, e.g., as described for bone lesions in [28]. However, this topic was beyond the scope of this study, and as no satisfying algorithms for total lesion segmentation in PSMA-PET are available, we decided for a manual segmentation not to alter our results by problems with a not fully validated segmentation process. Even as

manual segmentation shows several problems, e.g., interobserver variability, it is widely accepted in this context [19,29].

As one of the limitations of the current study, the gold standard, which is in our case, the human-experienced nuclear medicine physician, should be discussed. For the detection of pathological uptake, naturally, histopathology findings should be preferred. However, we had nearly 2500 hotspots in our training cohort of 72 patients, meaning nearly 35 hotspots per patient, including physiological uptake. Therefore, there is no option and justification for performing such a high number of biopsies. This is a general limitation of this kind of study.

Another general limitation that needs to be taken into account is that the whole study was performed on data of the same scanner. Although this was not part of the current study, in further steps, it should be analyzed how specific training and results are according to changes in scan devices or scan protocols. Such questions should be dealt with best in multicentric studies.

For further studies in ML analysis of medical image data, prostate cancer seems to be an ideal subject having a high and increasing prevalence. It should be easy to acquire "big data" and so to train, develop, and improve the algorithm. In the end, a large group of people would benefit from this study.

5. Conclusions

Computer-aided decision support systems analyzing radiomics features from pretherapeutic 68Ga-PSMA-PET-CT scans and leveraging state-of-the-art supervised machine learning methods have shown their significance to identify pathological uptake in patients with advanced prostate carcinoma even on unseen data. Further improvement, however, should be done in the identification of small glands and in cross-validation of different PET scanners and perhaps other centers to overcome limitations of interobserver variability and problems due to scan specifications.

Author Contributions: A.E.: Study concept, patient analysis, and writing the manuscript. S.M.: Study concept, data analysis and algorithm coding, and writing the manuscript. S.L.: Patient analysis and correcting the manuscript. M.E.: Study concept and correcting the manuscript. T.S.: Study concept, data analysis, and correcting the manuscript. R.A.B.: Study concept, lesion delineation, and correcting the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. All patients gave written and informed consent to the imaging procedure and for anonymized evaluation and publication of their data. Due to the retrospective character of the data analysis, an ethical statement was waived by the institutional ethical review board according to the professional regulations of the medical board of the state of Nordrhein-Westfalen, Germany.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available because, according to German "Datenschutz Grundverordnung 2016/679", patient data including images or part of images can only be used/seen within the hospital and by personnel of these hospitals if the patient has not agreed specifically to publish it. As this is a retrospective analysis, we have no agreement of the patients that we can make the data available openly, just that the data can be evaluated for studies.

Conflicts of Interest: R.A.B. is a Consultant for Bayer Healthcare (Leverkusen, Germany) and Eisai GmbH (Frankfurt, Germany). R.A.B. has a non-commercial research agreement and is on the speakers list with Mediso Medical Imaging (Budapest, Hungary). M.E. is a Consultant for Bayer Healthcare (Leverkusen, Germany) and Eisai GmbH (Frankfurt, Germany), IPSEN and Novartis. All other authors declare that there are no conflict of interest as well as consent for scientific analysis and publication.

Abbreviations

PC	prostate cancer
PSMA	prostate specific membrane antigen
PSA	prostate specific antigen
PET	positron emission tomography
CT	computed tomography
RFs	radiomics features
ML	machine learning
AI	artificial intelligence
AUC	area under the curve
SE	sensitivity
SP	specificity
VoI	volume of interest
gut	gastrointestinal tract

References

- 1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
- 2. Ferlay, J.; Lam, F.; Colombet, M.; Mery, L.; Pineros, M.; Znaor, A.; Soerjomataram, I. *Global Cancer Observatory: Cancer Today*; International Agency for Research on Cancer: Lyon, France, 2021.
- 3. Official Website of National Cancer Institute. Cancer Stat Facts: Prostate Cancer. 2019. Available online: https://seer.cancer.gov/ statfacts/html/prost.html (accessed on 12 April 2021).
- 4. Schwarzenboeck, S.M.; Rauscher, I.; Bluemel, C.; Fendler, W.P.; Rowe, S.P.; Pomper, M.G.; Afshar-Oromieh, A.; Herrmann, K.; Eiber, M. PSMA Ligands for PET Imaging of Prostate Cancer. J. Nucl. Med. 2017, 58, 1545–1552. [CrossRef] [PubMed]
- Wright, G.L.; Haley, C.; Beckett, M.L.; Schellhammer, P.F. Expression of prostate-specific membrane antigen in normal, benign, and malignant prostate tissues. Urol. Oncol. 1995, 1, 18–28. [CrossRef]
- Szabo, Z.; Mena, E.; Rowe, S.P.; Plyku, D.; Nidal, R.; Eisenberger, M.A.; Antonarakis, E.S.; Fan, H.; Dannals, R.F.; Chen, Y.; et al. Initial Evaluation of (18)FDCFPyL for Prostate-Specific Membrane Antigen (PSMA)-Targeted PET Imaging of Prostate Cancer. *Mol. Imaging Biol.* 2015, 17, 565–574. [CrossRef]
- Afshar-Oromieh, A.; Hetzheim, H.; Kratochwil, C.; Benesova, M.; Eder, M.; Neels, O.C.; Eisenhut, M.; Kübler, W.; Holland-Letz, T.; Giesel, F.L.; et al. The Theranostic PSMA Ligand PSMA-617 in the Diagnosis of Prostate Cancer by PET/CT: Biodistribution in Humans, Radiation Dosimetry, and First Evaluation of Tumor Lesions. J. Nucl. Med. 2015, 56, 1697–1705. [CrossRef] [PubMed]
- Fendler, W.P.; Calais, J.; Eiber, M.; Flavell, R.R.; Mishoe, A.; Feng, F.Y.; Nguyen, H.G.; Reiter, R.E.; Rettig, M.B.; Okamoto, S.; et al. Assessment of 68Ga-PSMA-11 PET Accuracy in Localizing Recurrent Prostate Cancer: A Prospective Single-Arm Clinical Trial. JAMA Oncol. 2019, 5, 856–863. [CrossRef]
- 9. Rauscher, I.; Maurer, T.; Fendler, W.P.; Sommer, W.H.; Schwaiger, M.; Eiber, M. (68)Ga-PSMA ligand PET/CT in patients with prostate cancer: How we review and report. *Cancer Imaging* **2016**, *16*, 14. [CrossRef]
- 10. Strack, C.; Seifert, R.; Kleesiek, J. Künstliche Intelligenz in der Hybridbildgebung. Radiologe 2020, 60, 405–412. [CrossRef]
- 11. Sollini, M.; Bartoli, F.; Marciano, A. Artificial intelligence and hybrid imaging: The best match for personalized medicine in oncology. *Eur. J. Hybrid. Imaging* **2020**, *4*, 24. [CrossRef]
- 12. Mayer-Schnberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work and Think*; John Murray Publishers: London, UK, 2014.
- 13. Goldenberg, S.L.; Nir, G.; Salcudean, S.E. A new era: Artificial intelligence and machine learning in prostate cancer. *Nat. Rev. Urol.* **2019**, *16*, 391–403. [CrossRef] [PubMed]
- 14. Hambarde, P.; Talbar, S.; Mahajan, A.; Chavan, S.; Thakur, M.; Sable, N. Prostate lesion segmentation in MR images using radiomics based deeply supervised U-Net. *Biocybern. Biomed. Eng.* **2020**, *40*, 1421–1435. [CrossRef]
- 15. Thrall, J.H.; Li, X.; Li, Q.; Cruz, C.; Do, S.; Dreyer, K.; Brink, J. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. J. Am. Coll. Radiol. **2018**, 15, 504–508. [CrossRef]
- Bundschuh, R.A.; Dinges, J.; Neumann, L.; Seyfried, M.; Zsótér, N.; Papp, L.; Rosenberg, R.; Becker, K.; Astner, S.T.; Henninger, M.; et al. Textural Parameters of Tumor Heterogeneity in ¹⁸F-FDG PET/CT for Therapy Response Assessment and Prognosis in Patients with Locally Advanced Rectal Cancer. J. Nucl. Med. 2014, 55, 891–897. [CrossRef] [PubMed]
- 17. Chicklore, S.; Goh, V.; Siddique, M.; Roy, A.; Marsden, P.K.; Cook, G.J.R. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur. J. Nucl. Med. Mol. Imaging* **2013**, *40*, 133–140. [CrossRef] [PubMed]
- Pyka, T.; Bundschuh, R.A.; Andratschke, N.; Mayer, B.; Specht, H.M.; Papp, L.; Zsótér, N.; Essler, M. Textural features in pre-treatment F18-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat. Oncol.* 2015, *10*, 100. [CrossRef] [PubMed]
- 19. Khurshid, Z.; Ahmadzadehfar, H.; Gaertner, F.C.; Papp, L.; Zsóter, N.; Essler, M.; Bundschuh, R.A. Role of textural heterogeneity parameters in patient selection for 177Lu-PSMA therapy via response prediction. *Oncotarget* **2018**, *9*, 33312–33321. [CrossRef]

- Moazemi, S.; Erle, A.; Lütje, S.; Gaertner, F.C.; Essler, M.; Bundschuh, R.A. Estimating the Potential of Radiomics Features and Radiomics Signature from Pretherapeutic PSMA-PET-CT Scans and Clinical Data for Prediction of Overall Survival When Treated with 177Lu-PSMA. *Diagnostics* 2021, 11, 186. [CrossRef]
- 21. Moazemi, S.; Khurshid, Z.; Erle, A.; Lütje, S.; Essler, M.; Schultz, T.; Bundschuh, R.A. Machine Learning Facilitates Hotspot Classification in PSMA-PET/CT with Nuclear Medicine Specialist Accuracy. *Diagnostics* **2020**, *10*, *622*. [CrossRef]
- 22. Official Company Website for the InterView FUSION: Software. Available online: https://www.mediso.de/Interview-fusion.html (accessed on 12 April 2021).
- 23. Kotsiantis, S.B. Supervised Machine Learning: A Review of Classification Techniques. Informatica 2007, 31, 249–268.
- 24. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- Simm, J.; de Abril, I.; Sugiyama, M. Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression. *Peer Rev.* 2014, E97-D, 1677–1681. [CrossRef]
- 26. Quinlan, J. Induction of Decision Trees. Mach. Learn. 1986, 1, 81–106. [CrossRef]
- 27. Moazemi, S.; Erle, A.; Khurshid, Z.; Lütje, S.; Muders, M.; Essler, M.; Schultz, T.; Bundschuh, R.A. Decision support for treatment with 177Lu-PSMA: Machine learning predicts response with high accuracy based on PSMA-PET/CT and clinical parameters. *Ann. Transl. Med.* **2021**, *9*, 818. [CrossRef]
- 28. Hammes, J.; Täger, P.; Drzezga, A. EBONI: A Tool for Automated Quantification of Bone Metastasis Load in PSMA PET/CT. J. Nucl. Med. 2018, 59, 1070–1075. [CrossRef] [PubMed]
- 29. Thuillier, P.; Liberini, V.; Rampado, O.; Gallio, E.; De Santi, B.; Ceci, F.; Metovic, J.; Papotti, M.; Volante, M.; Molinari, F.; et al. Diagnostic Value of Conventional PET Parameters and Radiomic Features Extracted from 18F-FDG-PET/CT for Histologic Subtype Classification and Characterization of Lung Neuroendocrine Neoplasms. *Biomedicines* **2021**, *9*, 281. [CrossRef] [PubMed]

3. Danksagung

Mein Dank gilt zunächst meinem Doktorvater Prof. Dr. Dr. Bundschuh für die hervorragende Betreuung, Hilfe und Unterstützung. Ich bedanke mich ganz besonders für die gute Einarbeitung, die schnelle und zuverlässige Kommunikation und die strukturierte Anleitung.

Ich bedanke mich bei Sobhan Moazemi für die unkomplizierte Zusammenarbeit bei diesem Projekt, den interessanten Austausch und die Beratung bei allen Rückfragen, außerdem für die tolle Unterstützung bezüglich der Veröffentlichung unseres Papers und den damit verbundenen organisatorischen Aspekten.

Ich bedanke mich auch bei Susanne Lütje für die Unterstützung bei organisatorischen und administrativen Problemen.