

Aus dem Institut für Pathologie
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Multimodales Deep Learning zur Prognosevorhersage für das klarzellige
Nierenzellkarzinom

Inauguraldissertation
zur Erlangung des Doktorgrades der
Medizin
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Stefan Jonathan Klaus Schulz
aus Wiesbaden

Mainz, 2025

Wissenschaftlicher Vorstand: Univ.-Prof. Dr. med. Philipp Drees

Tag der Promotion: 04. Dezember 2025

Nachnutzungslizenz: Namensnennung (CC-BY-4.0)

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	I
1 Zusammenfassung	1
2 Originalarbeit	4
3 Abschließende Diskussion	5
3.1 Plausibilität der Anwendung von Deep Learning auf multimodale medizinische Bilddaten des klarzelligen Nierenzellkarzinoms	5
3.2 Anwendung und Nutzen von Verfahren zur Steigerung der Erklärbarkeit Künstlicher Intelligenz	6
3.3 Prognostische Bedeutung von Genmutationen und Genexpressionssignaturen für das klarzellige Nierenzellkarzinom	7
3.4 Ansätze zur Verbesserung von Deep Learning-Modellen mittels Schwarmlernen und der Erzeugung synthetischer Bilddaten.....	8
4 Literaturverzeichnis.....	9
5 Anhang: Publierte Originalarbeit	13
6 Danksagung	23
7 Tabellarischer Lebenslauf	24

Abkürzungsverzeichnis

5YSS / 5-JÜR	5-year survival status / 5-Jahres-Überlebensrate
AI	Artificial intelligence
BAP1	BRCA1-associated protein-1
BRCA1	Breast Cancer gene 1
ccRCC	Clear cell renal cell carcinoma
CD4 / CD8	Cluster of differentiation 4 / 8
CNN	Convolutional Neural Network
CT	Computertomographie
CTLA-4	Cytotoxic T-lymphocyte-associated protein 4
DL	Deep Learning
HIF1 α	Hypoxia-inducible factor 1-alpha
H&E	Hämatoxylin-Eosin
IMDC	International Metastatic Renal cell Carcinoma Database Consortium
ISUP	International Society of Urological Pathology
KI	Künstliche Intelligenz
KIRC	Kidney Renal Clear cell carcinoma
MMDLM	Multimodales Deep Learning Modell
MRT	Magnetresonanztomographie
mTOR	Mechanistic/mammalian target of rapamycin
NZK	Nierenzellkarzinom
PBRM1	Polybromo1(-Gen)
PD-1 / PD-L1	Programmed death-1 / PD-1 receptor ligand
TAM	Tumor-associated macrophage
TME	Tumor microenvironment
TML	Tumormutationslast
VEGFR	Vascular endothelial growth factor receptor
VHL	Von-Hippel-Lindau(-Tumorsuppressorgen)
WSI	Whole-slide image

1 Zusammenfassung

Das Nierenzellkarzinom (NZK) stellt mit einem Anteil von 2-3% (Bray et al. 2017) an allen malignen Neoplasien des Menschen eine vergleichsweise seltene Krebsentität dar. Allerdings geht der mit etwa 70-75% häufigste Subtyp des NZK – das klarzellige Nierenzellkarzinom (ccRCC bzw. KIRC) – mit einer relevanten Morbidität und Mortalität einher, wenngleich die Anwendung von (VEGFR-)Tyrosinkinase-Inhibitoren, Multikinase-Inhibitoren und Immuncheckpoint-Inhibitoren das mittlere Überleben der Patient*innen über den Zeitraum 2000-2020 bereits deutlich verbessern konnte (Demasure et al. 2022). NZK sind in den frühen klinischen Stadien häufig ohne Symptome (Vasudev et al. 2020). Dies führt dazu, dass zum Zeitpunkt des Auftretens der charakteristischen Symptomtrias (Makrohämaturie, Flankenschmerz und palpabler abdomineller Raumforderung) oder dem Auftreten paraneoplastischer Syndrome bereits bei etwa einem Viertel der Patient*innen eine primäre Metastasierung vorliegt (Chen et al. 2021, Dudani et al. 2020). Daneben finden sich häufig Lymphknotenmetastasen und zum Teil Metastasen in Nachbarorganen wie der Nebenniere oder dem Pankreas (Dudani et al. 2020). Zudem ist aufgrund der Latenz zwischen Tumorentstehung und Diagnosestellung die Wahrscheinlichkeit für einen Progress der Primärtumorgröße (T-Stadium) sowie für das Vorliegen einer, durch Akkumulation weiterer Mutationen (Pang et al. 2018) schlechter differenzierten Tumorzellpopulation mit aggressiverem Wachstums- und Metastasierungverhalten, erhöht. Es stehen bislang im Gegensatz zu anderen Tumorerkrankungen wie etwa dem Bronchial-, Mamma- oder dem kolorektalen Karzinom noch keine ausreichend etablierten Biomarker zur Risikostratifizierung und Therapieleitung des NZK zur Verfügung. So sind z.B. weder der PD-L1-Status noch die Tumormutationslast (TML) für die Therapie des NZK prädiktiv. Womöglich wird erst die Integration komplexer molekularer und histologischer Datenmodalitäten unter Abbildung der individuellen Tumorerheterogenität und –immunogenität ausreichend spezifische therapieleitende Biomarker des NZK hervorbringen (Saliby et al. 2024). Gemäß der aktuellen S3-Leitlinie sollen Therapieentscheidungen zunächst weiterhin basierend auf histopathologischer Subentität nach aktueller WHO-Klassifikation getroffen werden, wobei folgende Parameter ergänzend betrachtet werden sollen: Sonderformen gemäß Vancouver-Klassifikation, TNM-Staging, Bestimmung des Tumorgrads nach Fuhrmann bzw. nach WHO-ISUP, abdominelles CT/MRT-Staging, ggf. Ganzkörper low-dose CT bei Verdacht auf ossäre sowie Schädel-MRT bei Verdacht auf kraniale Metastasierung (DGHO & DGU LL4.0: 2023, Kristiansen et al. 2015). Die Prognoseabschätzung fußt hierbei auf Nomogrammen, welche anhand klinischer Parameter wie dem Karnofsky-Performancessstatus, der zeitlichen Abfolge und Lokalisation der Metastasenbildung oder laborchemischen Konstellationen erstellt wurden (DGHO & DGU LL4.0: 2023). Diese Faktoren bestimmen zusammen mit dem Zeitintervall zwischen Diagnosestellung und Beginn einer Systemtherapie als IMDC-Kriterien die Einstufung des/der Patient*in in ein günstiges, intermediäres oder ungünstiges Risikoprofil. Das mediane Gesamtüberleben entspricht in diesen Risikogruppen etwa 43, 22 bzw. 8 Monate. Hiervon hängt in der Folge auch die Wahl der Erstlinientherapie bei fortgeschrittenem/metastasiertem ccRCC ab (DGHO & DGU LL4.0: 2023, Heng et al. 2013).

Die vorliegende Arbeit zielt daher darauf ab anhand multimodaler, d.h. sowohl histopathologischer als auch radiologischer Bilddaten sowie Genmutationsdaten ein Deep Learning (DL)-Modell zu generieren, welches die individuelle Prognose einer Patient*in abschätzen kann. Eine verbesserte Prognosevorhersage, resultierend aus der Verwendung möglichst umfassender Datenbestände aus der urologischen, radiologischen und (molekular-)pathologischen Diagnostik kann zur weiteren Individualisierung der

Empfehlungen hinsichtlich der Therapie- und Nachsorgeschemata führen und damit das klinische Outcome von Patient*innen mit klarzelligen Nierenzellkarzinomen verbessern. Deep Learning-Ansätze finden aktuell immer stärkere Anwendung im Bereich medizinischer Fragestellungen, ins Besondere hinsichtlich radiologischer und histopathologischer Bilddaten (Anaya-Isaza et al. 2021). Neben Aufgabenstellungen der Bildsegmentierung (Ranjbarzadeh et al. 2021) oder der Tumorklassifikation (Byeon et al. 2022, Iizuka et al. 2020), kann Deep Learning hinsichtlich vieler weiterer relevanter Zielparameter trainiert werden, wie bspw. zur Vorhersage der Mikrosatelliten-Instabilität des Magen- und kolorektalen Karzinoms (Kather et al. 2019) oder mit dem Ziel der Prognosevorhersage anhand individueller Patientendaten (Foersch et al. 2023, Vale-Silva et al. 2021).

Zur Durchführung der vorliegenden Studie wurden zwei ccRCC-Patientenkohorten mit entsprechenden histopathologischen, radiologischen (CT/MRT-Abdomen), Überlebens- und Genmutationsdaten (Exom-Sequenzierung) verwendet. Einerseits wurden Daten zu 230 Fällen von der TCGA-Datenbank (Tomczak et al. 2015) bezogen, andererseits diente eine Kohorte von 18 Mainzer Patient*innen als externe Testdatenmenge zur Evaluation der generierten Modelle. Beide Kohorten umfassten sowohl nicht-metastasierte als auch metastasierte/fortgeschrittene ccRCC-Fälle. Für jeden diese/r Patient*innen wurden zunächst unimodale DL-Modelle auf den jeweiligen Datenmodalitäten trainiert. Im Fall der histologischen Schnittbilder wurden zwei unterschiedlich stark vergrößerte Bildebenen verwendet, um das Erlernen zellulärer als auch größerer, zusammenhängender Gewebekontext-Merkmale zu ermöglichen. Die Annotation von Tumorgewebe wurde mittels der QuPath-Bildauswertungssoftware (Bankhead et al. 2017) vorgenommen und Bildkacheln („Tiles“) aus ccRCC-Arealen wurden exportiert. Hiernach erfolgte die Normalisierung (Vahadane et al. 2015) und Augmentierung, d.h. die Datenmenge für das Training wurde durch verschiedene Bildoperationen erhöht. Für die bezogenen abdominellen CT/MRT-Datensätze wurden pro Patient*in drei radiologische Bildkacheln mit der jeweils größten Tumorausdehnung in der sagittalen, koronalen/frontalen und axialen Ebene ausgewählt und exportiert, um das Tumolvolumen abzubilden. Nach Abschluss des unimodalen Trainings wurde basierend auf den vortrainierten Gewichten innerhalb der unimodalen Modelle ein multimodales DL-Modell (MMDLM) trainiert. Hierbei wurde wobei basierend auf einer modifizierten Version des „Pathomic Fusion“-Ansatzes von Chen et al. (2019) der jeweilige Beitrag der unimodalen Modelle gewichtet und diese anschließend miteinander zu einem multimodalen Tensor verrechnet. Zielparameter des Trainings der unimodalen wie des multimodalen Modells war die Vorhersage des krankheitsbezogenen Überlebens einerseits in Form des 5-Jahres-Überlebensstatus (5YSS) sowie in Form des Harrell'schen Konkordanz-Index (C-Index), welcher die Güte der Vorhersage der Sortierung der Patient*innen entlang ihrer Überlebenszeiträume bewertet.

Bei Auswertung der Ergebnisse zeigte sich im Rahmen einer multivariaten Cox-Regression die Prognosevorhersage des MMDLM als unabhängiger Risikofaktor mit einer Hazard Ratio von 3,9 und übertraf dabei klinische Parameter wie den Differenzierungsgrad sowie den Fernmetastasen- und regionären Lymphknotenmetastasierungsstatus. Die zusätzliche Verwendung der zehn häufigsten ccRCC-Genmutationen im Rahmen des multimodalen Trainings zeigte keinen signifikanten Unterschied im Vergleich mit dem ausschließlich auf radiologischen und multiskaliert-histopathologischen Bilddaten trainierten MMDLM. Insgesamt erreichte das MMDLM bei der Evaluation anhand der Mainzer Testdatenkohorte eine Genauigkeit (Accuracy) von $79,17\% \pm 9,8\%$ mit einem Maximum von $94,44\%$. Zudem zeigte das MMDLM eine signifikant bessere Vorhersagekraft im Vergleich zu auf

radiologischen respektive auf histologischen Bilddaten geringerer Vergrößerung trainierten unimodalen Modellen und zeigte bei Betrachtung entsprechender Kaplan-Meier-Kurven eine klare und statistisch signifikante Separation zwischen Patient*innenpopulationen mit hohem vs. niedrigem Risiko bzgl. des erkrankungsbezogenen Überlebens (Schulz et al. 2021).

2 Originalarbeit

Ergebnisse der vorliegenden Arbeit (Schulz et al. 2021) wurden am 24. November 2021 unter dem Titel „Multimodal Deep Learning for Prognosis Prediction in Renal Cancer“ bei Frontiers in Oncology veröffentlicht:

Schulz S, Woerl A-C, Jungmann F, Glasner C, Stenzel P, Strobl S, Fernandez A, Wagner D-C, Haferkamp A, Mildemberger P, Roth W and Foersch S. **Multimodal Deep Learning for Prognosis Prediction in Renal Cancer.** *Front. Oncol.* 11:788740. doi:10.3389/fonc.2021.788740 (2021).

3 Abschließende Diskussion

Mithilfe von multimodalen Deep Learning-Ansätzen kann die Prognosevorhersage für Patient*innen mit klarzelligem Nierenzellkarzinom im Vergleich zu unimodal trainierten Modellen verbessert werden, wodurch sich perspektivisch eine individuellere Therapieempfehlung mit Verbesserung des tumorbezogenen Überlebens erzielen lassen könnte. Im Folgenden sollen die Plausibilität der Verwendung von Deep Learning für die beschriebene Aufgabenstellung sowie Limitationen des in der vorliegenden Studie verwendeten Datenmaterials diskutiert werden. Zudem wird der Nutzen von Verfahren zur Steigerung der Transparenz von DL-Modellen dargestellt. Zuletzt erfolgt ein Ausblick auf aktuell in der Entwicklung befindliche neuartige Möglichkeiten zur Verbesserung des DL-Trainings durch sog. „Swarm Learning“-Ansätze.

3.1 Plausibilität der Anwendung von Deep Learning auf multimodale medizinische Bilddaten des klarzelligen Nierenzellkarzinoms

Die Anwendung eines Deep Learning-Ansatzes zur Prognosevorhersage im Falle des klarzelligen NZK mittels histopathologischer Bilddaten erscheint insofern aussichtsreich, als für diese Entität eine enge Phänotyp-Prognose-Relation besteht. Dies konnte für eine Reihe morphologischer Kriterien gezeigt werden. Beispiele hierfür bilden sowohl eine mit einer schlechteren Prognose zunehmende Eosinophilie des besonders anfänglich oft namensgebend klar- bzw. transparent imponierenden Glykogen- und Neutralfett-haltigen ccRCC-Zytoplasmas (Nilsson et al. 2020, Ohe et al. 2023). Auch das Ausmaß der Tumorangiogenese, das Auftreten von Tumornekrosearealen sowie von sarkomatoiden oder rhabdoiden Wachstumsmustern (Zhang 2018 et al., Blum et al. 2020, Mohamed et al. 2020) sind mit einer ungünstigen Prognose assoziiert. Hinsichtlich der Wachstumsmuster wurde eine Entwicklung des ccRCC aus zystischen Vorläuferläsionen postuliert (Moch et al. 2010, Montani et al. 2010), sodass das Erkennen zystischer oder tubulozystischer Anteile als mögliche Korrelate eines eher frühen Tumorstadiums ebenfalls zur Prognosevorhersage beitragen könnte. Die bisher beschriebenen morphologischen Aspekte lassen sich bereits in überblicksartigen, schwach vergrößerten Bildausschnitten erkennen und mutmaßlich durch das DL-Modell erlernen. Werden hingegen stärker vergrößerte Ebenen des Whole-Slide Images (WSI), als zusätzliche Modalität verwendet, könnte dies dem DL-Modell ermöglichen Details zellulärer Strukturen und kleinerer Zellverbände zu erlernen. Beispielsweise könnte dies das Erlernen von Zellkernatypien wie Kernpleomorphien, (atypischen) Mitosefiguren oder prominenten Nukleolen ermöglichen, also von Merkmalen, welche bereits Bestandteil bisher diagnostisch eingesetzter histopathologischer Scores wie der Fuhrman-Gradierung (Ficarra et al. 2005) oder des seit 2016 durch die WHO und ISUP etablierten ISUP-Scores sind. Auch das Erkennen von Veränderungen der Tumor-Mikroumgebung (TME), v.a. im Hinblick auf Anzahl und Verteilung zytotoxischer T-Lymphozyten könnte die individuelle Prognosevorhersage noch weiter verbessern (Vuong et al. 2019, Giraldo et al. 2017). Beispielsweise konnte gezeigt werden, dass das Therapieansprechen von Patient*innen mit Nachweis sarkomatoider Tumoranteile, welche mit einer höheren PD-L1-Expression und einer geringer ausgeprägten Angiogenese-Signatur im Vergleich zu nicht-sarkomatoid-differenzierten ccRCC einhergehen (McDermott et al. 2019) durch eine spezifische Kombinationstherapie mit Ipilimumab (CTLA-4-Inhibitor) und Nivolumab (PD-1-Inhibitor) deutlich gesteigert werden konnte (Rini et al. 2020).

Analog sind auch für radiologische Bilddaten prognostisch ungünstige Bildmorphologien bekannt (Abou Elkassem et al. 2021). Hierzu zählen Tumorthromben in den Nierenvenen oder der Vena cava inferior, vergrößerte Lymphknoten oder eine Metastasierung per continuitatem in anatomisch benachbarte Strukturen. Des Weiteren können radiologische Korrelate von Tumornekrosen, Verkalkungen und Septenbildung sowie eine Kontrastmittel-Anreicherung oder Inhomogenität der Nierenläsion einem DL-Modell als erlernbare Merkmale dienen. Eine Abwesenheit solcher Kriterien deutet bei runden, kleinen (< 3 cm) und gut abgrenzbaren Veränderungen auf die Differentialdiagnose einer (ggf. auch nebenbefundlich vorhandenen) zystischen Bosniak-1/-2 Läsion hin (Israel et al. 2005). Neben diesen, bereits durch Radiolog*innen verwendeten Anhaltspunkten, ist es im Sinne des neuen Forschungsfelds der „Radiomics“ denkbar, dass bestimmte Veränderungen der menschlichen visuellen Sinnesphysiologie nicht zugänglich sind. Diese können jedoch durch KI-Ansätze erkannt und als Bildmerkmale erlernt werden (Lambin et al. 2017). Ein Indiz für diese These besteht etwa in ersten vielversprechenden Studien zur allein auf CT/MRT-Bilddaten basierenden Klassifikation von Nierenläsionen mit mehrheitlich guter Abgrenzung des NZK von benignen Nierentumoren wie dem Angiomyolipom oder dem Onkozytom, welche perspektivisch die Rate operativer Übertherapie im Sinne von (Teil-)Nephrektomien mit anschließender benigner histologischer Wertung verringern könnte (Mühlbauer et al. 2021).

3.2 Anwendung und Nutzen von Verfahren zur Steigerung der Erklärbarkeit Künstlicher Intelligenz

Aufgrund der zuvor beschriebenen Zusammenhänge erscheint es plausibel, dass insbesondere im Falle des NZK bekannte morphologische Korrelate der Prognose des Patienten bestehen, welche einem Deep Learning-Modell als erlernbare Merkmale dienen können. Darüber hinaus besteht die Möglichkeit mittels sog. „explainable AI“-Verfahren, also im Sinne „transparenter KI“, sowohl die Gewichtung bereits etablierter morphologischer Kriterien zu überprüfen als auch potentiell bislang unbekannte morphologische Korrelate der Gradierung oder der Prognosevorhersage zu entdecken. Auch lässt sich durch die Anwendung von Verfahren zur Steigerung der Transparenz überprüfen, ob plausible Merkmale („Features“) der Bilddaten erfasst werden und nicht etwa eine zu starke Anpassung („Overfitting“) an die spezifischen Trainingsdaten vorliegt. Beispielsweise könnten Erkennungsmerkmale der Objektträgerscanner oder der Computertomographen verschiedener Klinika mit unterschiedlicher Prognosekorrelation in Abhängigkeit des Standorts bestehen, was im Falle eines „Overfitting“ an diesen Aspekt der Trainingsdaten die Generalisierbarkeit des Modells beeinträchtigt.

In der vorliegenden Studie wurden sog. Class activation maps (CAMs) angewendet, welche die Überlagerung z.B. einer histopathologischen Bildkachel mit einer farbcodierten Darstellung der Gewichtung bestimmter Bildareale bei der Prognosestellung durch das Neuronale Netzwerk erlauben (Zhou et al. 2015). Hierbei konnten die in der Literatur beschriebenen prognostisch relevanten Eigenschaften bestätigt werden. Hierzu gehörte einerseits die Assoziation eines klarzelligen Zytoplasmaaspekts oder papillärer Wachstumsmuster mit einem niedrigeren Risiko sowie andererseits eine Assoziation von vermehrter Kapillarisation des Tumorgewebes mit einem höheren Risikostatus. Ein alternativer Ansatz bestünde in der Nutzung sog. „Occlusion Sensitivity Maps“ (Huff et al. 2021). Hierbei werden verschiedene Areale einer Bildkachel verdeckt und anschließend

geprüft, wie stark die Fähigkeit der Klassifikation durch Verlust dieses Bildareals abfällt, um dessen Beitrag zur Vorhersage zu bemessen.

In der vorliegenden Arbeit wurde als weitere Möglichkeit die Plausibilität der Klassifikation / Vorhersage durch das DL-Modell zu überprüfen ein stichprobenartiges „Markup“ von WSIs durchgeführt, wobei ein gesamtes Bild eines Gewebsschnittes / WSI mit farbkodierten einzelnen Bildkacheln überlagert wird. Hierdurch lässt sich überblicksartig das Ausmaß der Übereinstimmung der Kachel-basierten Urteile zu einem Patientenfall hinsichtlich ihrer prognostischen Zuordnung qualitativ einschätzen.

3.3 Prognostische Bedeutung von Genmutationen und Genexpressionssignaturen für das klarzellige Nierenzellkarzinom

In einem weiteren multimodalen Ansatz verwendeten wir neben den radiologischen und histopathologischen Bilddaten auch Genmutationen als zusätzliche Datenmodalität. Um dieses erweiterte MMDLM nicht zu komplex zu gestalten wurden hierbei die zehn häufigsten Genmutationen zusammen mit Daten zu Kopienzahlveränderungen (copy number alterations) des ccRCC ausgewählt und über eine vollständig verknüpfte Schicht des Neuronalen Netzwerks („Fully connected layer“ bzw. „Dense layer“) in das bestehende multimodale Modell eingefügt. Interessanterweise zeigte sich durch das Einbringen dieser Daten keine signifikante Verbesserung der Vorhersage durch das MMDLM. Dies ist möglicherweise auf die Tatsache zurückzuführen, dass in der verwendeten Patientenkohorte kein statistisch signifikanter Zusammenhang zwischen den einzelnen Genmutationen bzw. Kopienzahlveränderungen und der Prognose bestand (Schulz et al. 2021), wenngleich in der Literatur bei Auswertungen größerer Patientenkohorten einzelne prognostisch relevante Genmutationen (z.B. BAP1) bzw. prognostisch bedeutsame Genmutationskombinationen (z.B. [VHL + PRBM1]) spezifisch für das ccRCC identifiziert werden konnten (Pang et al. 2018, Ricketts et al. 2018). Es ist denkbar, dass eine Erweiterung des Datenbestandes etwa durch Einbeziehen von DNA-Methylierungs- und differentiellen Genexpressionsdaten die Vorhersagekraft eines multimodal trainierten Netzwerks steigern könnte. Aufgrund kleiner Studienpopulationen wurde bis dato allerdings nur für wenige Panels von Methylierungsmarkern eine prognostische Signifikanz beschrieben für welche eine Gültigkeit über verschiedene Kohorten hinweg gezeigt werden konnte (Lasseigne et al. 2018). Als vielversprechende Kandidaten erscheinen derzeit Genexpressionsmuster verschiedener an Hypoxie-Signalfaden beteiligter Gene innerhalb der ccRCC-Tumorzellen. So konnte z.B. die Kombination von sechs solcher Gene zu einem Hypoxie-bezogenen Risiko-Score bisherige klinische Nomogramme des ccRCC verbessern (Li et al. 2021). Zhong et al. (2021) identifizierten darüber hinaus anhand einer Kombination aus Genexpressions-, somatischen Mutations- und Kopienzahlvariationsdaten zwei distinkte molekulare Hypoxie- und Tumorummunogenität-bezogene ccRCC-Subtypen, vermittels derer die Identifikation von Hochrisikopatienten oder die Vorhersage des Therapieansprechens auf Immuntherapien verbessert werden könnte. Long et al. (2021) konnten für das ccRCC einen prognostisch relevanten Zusammenhang zwischen einer JAK3-Promoterhypomethylierung und einer resultierend gesteigerten JAK3-Genexpression mit der Infiltration von CD4- und CD8-positiven T-Lymphozyten in das Tumorgewebe bei gleichzeitig prognostisch ungünstig erhöhter Immunsuppression durch gesteigerte Genexpression von Immuncheckpointmolekülen nachweisen.

Diese beispielhaft beschriebenen Marker werden aktuell jedoch noch nicht als Bestandteil der Standarddiagnostik erhoben, sodass sie derzeit noch nicht zur weiteren Verbesserung von Deep Learning-Modellen verwendet werden können, welche auf eine ausreichende Anzahl möglichst vollständiger Patientendatensätze angewiesen sind.

3.4 Ansätze zur Verbesserung von Deep Learning-Modellen mittels Schwarmlernen und der Erzeugung synthetischer Bilddaten

Bei der Anwendung von Deep Learning auf medizinische Bilddaten wird oft auf öffentlich verfügbare Datenbanken in Verbindung mit lokalen Datenbeständen der jeweiligen Institution zurückgegriffen, wobei je nach Entität häufig insgesamt nur wenige Hundert Fälle verfügbar sind. Die Performanz der Deep Learning-Modelle könnte jedoch i.d.R. mit steigender Ausgangsdatenmenge weiter gesteigert werden (Alom et al. 2019). Daher reichen die an einzelnen Einrichtungen verfügbaren Datenbestände oft nur zur Erbringung von orientierenden Machbarkeitsnachweisen, ohne zu klinisch anwendbaren Modellen zu führen. Ein Austausch von medizinischen (Roh-)Bilddaten selbst in pseudonymisierter oder anonymisierter Form zwischen Forscher*innen verschiedener Kliniken oder Institutionen ist aufgrund datenschutzrechtlicher Regularien oft nicht möglich. In diesem Zusammenhang können Ansätze des Schwarmlernens („Swarm learning“) in Form einer wiederholten dezentralen Fusion der erlernten Parameter des Neuronalen Netzwerks ohne Notwendigkeit des Austauschs der eigentlichen Bilddaten eine Lösung bieten, um kooperativ zu leistungsstarken Modellen zu gelangen ohne sensible Daten preiszugeben (Warnat-Herresthal et al. 2021). Saldanha et al. (2022) konnten am Beispiel des kolorektalen Karzinoms die Umsetzung eines solchen Schwarmlernen-Ansatzes basierend auf histopathologischen Daten sowie die Überlegenheit gegenüber rein lokal trainierten Modellen demonstrieren.

Ein weiterer Ansatz zur Erhöhung der Datenmenge für das Training von DL-Modellen besteht in der Erzeugung von synthetischen Daten ausgehend von realen Bilddaten. Zur Umsetzung stehen derzeit verschiedene Ansätze wie etwa erzeugend-gegnerische Netzwerke („Generative-Adversarial Networks“) (Karras et al. 2021), Diffusionsmodelle (Dhariwal et al. 2021) oder Variations-Autoencoder (Oord et al. 2017) zur Auswahl. Derzeit beschäftigt sich ein BMBF-gefördertes Konsortium unter Beteiligung der Universitätsklinika Aachen, Dresden, Heidelberg, Würzburg und Mainz mit der Implementierung generativer Modelle zur Erzeugung synthetischer Daten in Verbindung mit Schwarmlernen, um vielfältige radiologische wie histopathologische Klassifizierungs- und Prognoseproblemstellungen bezüglich des ccRCC mittels Deep Learning zu bearbeiten (s. Quelle ‚BMBF‘).

4 Literaturverzeichnis

Abou Elkassem AM, Lo SS, Gunn AJ, Shuch BM, Dewitt-Foy ME, Abouassaly R, Vaidya SS, Clark JI, Louie AV, Siva S, Grosu AL, Smith AD. **Role of Imaging in Renal Cell Carcinoma: A Multidisciplinary Perspective.** *Radiographics* 2021 Sep-Oct;41(5):1387-1407. doi: 10.1148/rg.2021200202. Epub 2021 Jul 16. PMID: 34270355 (2021).

Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, Van Essen BC, Awwal AAS, Asari VK. **A State-of-the-Art Survey on Deep Learning Theory and Architectures.** *Electronics*. 2019; 8(3):292. <https://doi.org/10.3390/electronics8030292> (2019)

Anaya-Isaza A, Mera-Jiménez L, Zequera-Diaz M. **An overview of deep learning in medical imaging.** *Informatics in Medicine Unlocked, Volume 26, 2021, 100723, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2021.100723> (2021).*

Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW. **QuPath: Open Source Software for Digital Pathology Image Analysis.** *Sci Rep* (2017) 7(1):16878. doi: 10.1038/s41598-017-17204-5 (2017).

Blum KA, Gupta S, Tickoo SK, Chan TA, Russo P, Motzer RJ, Karam JA, Hakimi AA. **Sarcomatoid renal cell carcinoma: biology, natural history and management.** *Nat Rev Urol*. 2020 Dec;17(12):659-678. doi: 10.1038/s41585-020-00382-9. Epub 2020 Oct 13. PMID: 33051619; PMCID: PMC7551522 (2020).

BMBF, SWAG-Verbund: Baeßler B, Truhn D, Försch S, Engelhardt S, Kather J. [Internet] <https://www.gesundheitsforschung-bmbf.de/de/swag-schwamlernen-und-generative-modelle-zur-synthese-und-nutzbarmachung-hochqualitativer-15932.php>; abgerufen am 02.03.2023

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. **Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.** *CA Cancer J Clin* 68 (6):394–424. doi: 10.3322/caac.21492 (2018).

Byeon S-J, Park J, Cho YA, Cho B-J. **Automated histological classification for digital pathology images of colonoscopy specimen via deep learning.** *Sci Rep* 12, 12804 (2022). <https://doi.org/10.1038/s41598-022-16885-x> (2022).

Chen RJ, Ming YL, Wang J, Williamson DFK, Rodig SJ, Lineman NI, Mahmood F. **Pathomic Fusion: An Integrated Framework for fusing histopathology and genomic features for cancer diagnosis and prognosis.** *IEEE Trans Med Imaging* (2020), 1–1. doi: 10.1109/TMI.2020.3021387 (2020).

Chen SH, Xu LY, Wu YP, KE Z-B, Huang P, Lin F, Li X-D, Xue X-Y, Wei Y, Zheng Q-S, Xi N. **Tumor volume: a new prognostic factor of oncological outcome of localized clear cell renal cell carcinoma.** *BMC Cancer* 21, 79 (2021). <https://doi.org/10.1186/s12885-021-07795-8> (2021).

Demasure S, Spriet I, Debruyne PR, Laenen A, Wynendaele W, Baldewijns M, Dumez H, Clement PM, Wildiers H, Schöffski P, Roussel E, Kinget L, Albersen M, Beuselinc B. **Overall survival improvement in patients with metastatic clear-cell renal cell carcinoma between 2000 and 2020: a retrospective cohort study.** *Acta Oncologica*, 61:1, 22-29, DOI: 10.1080/0284186X.2021.1989720 (2022)

DGHO & DGU, S3-Leitlinie Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms (Version 4.0), Februar 2023. Verfügbar unter: <https://www.leitlinienprogramm-onkologie.de/leitlinien/nierenzellkarzinom/>, Zugriff am 02.03.2023. (2023)

Dhariwal P, Nichol A. **Diffusion Models Beat GANs on Image Synthesis** [Internet]. *arXiv [cs.LG]*. 2021. Available from: <http://arxiv.org/abs/2105.05233> (2021).

Dorević G, Matusan-Ilijas K, Babarović E, Hadzisejdić I, Grahovac M, Grahovac B, Jonjić N. **Hypoxia inducible factor-1alpha correlates with vascular endothelial growth factor A and C indicating worse prognosis in clear cell renal cell carcinoma.** *J Exp Clin Cancer Res*. 2009 Mar 20;28(1):40. doi: 10.1186/1756-9966-28-40. PMID: 19302703; PMCID: PMC2664792 (2009).

Dudani S, De Velasco G, Wells C, Gan CL, Donskov F, Porta C, Fraccon A, Pasini F, Hansen AR, Bjarnason GA, Beuselink B, Pal SK, Hotte SJ, Lalani A-KA, Yuasa T, Kanesvaran R, Reaume MN, Canil CM, Choueiri TK, and Heng DY. **Sites of metastasis and survival in metastatic renal cell carcinoma (mRCC): Results from the International mRCC Database Consortium (IMDC).** *Journal of Clinical Oncology* 2020 38:6_suppl, 642-642 (2020).

Ficarra V, Martignoni G, Maffei N, Brunelli M, Novara G, Zanolla L, Pea M, Artibani W. **Original and reviewed nuclear grading according to the Fuhrman system: a multivariate analysis of 388 patients with conventional renal cell carcinoma.** *Cancer.* 2005 Jan 1;103(1):68-75. doi: 10.1002/cncr.20749. PMID: 15573369 (2005).

Finley DS, Pantuck AJ, Belldegrun AS. **Tumor biology and prognostic factors in renal cell carcinoma.** *Oncologist.* 2011;16 doi: 10.1634/theoncologist.2011-S2-04 (2011).

Foersch S, Glasner C, Woerl AC, Eckstein M, Wagner DC, Schulz S, Kellers F, Fernandez A, Tserea K, Kloth M, Hartmann A, Heintz A, Weichert W, Roth W, Geppert C, Kather JN, Jesinghaus M. **Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer.** *Nat Med.* 2023 Feb;29(2):430-439. doi: 10.1038/s41591-022-02134-1. Epub 2023 Jan 9. PMID: 36624314 (2023).

Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, Sanchez-Salas R, Ingels A, Oudard S, Moatti A, Buttard B, Bourass S, Germain C, Cathelineau X, Fridman WH, Sautès-Fridman C. **Tumor-Infiltrating and Peripheral Blood T-cell Immunophenotypes Predict Early Relapse in Localized Clear Cell Renal Cell Carcinoma.** *Clin Cancer Res* 1 August 2017; 23 (15): 4416–4428. <https://doi.org/10.1158/1078-0432.CCR-16-2848> (2017).

Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, Mackenzie M, Wood L, Donskov F, Tan MH, Rha SY, Agarwal N, Kollmannsberger C, Rini BI, Choueiri TK. **External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study.** *Lancet Oncol.* 2013 Feb;14(2):141-8. doi: 10.1016/S1470-2045(12)70559-4. Epub 2013 Jan 9. PMID: 23312463; PMCID: PMC4144042 (2013).

Huff DT, Weisman AJ, Jeraj R. **Interpretation and visualization techniques for deep learning models in medical imaging.** *Phys Med Biol.* 2021 Feb 2;66(4):04TR01. doi: 10.1088/1361-6560/abcd17. PMID: 33227719; PMCID: PMC8236074 (2021).

Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. **Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours.** *Sci Rep* 10, 1504 (2020). <https://doi.org/10.1038/s41598-020-58467-9> (2020).

Israel GM, Bosniak MA. **How I do it: evaluating renal masses.** *Radiology* 2005;236(2):441-450 (2005).

Karras T, Aittala M, Laine S, Härkönen E, Hellsten J, Lehtinen J, Aila T. **Alias-Free Generative Adversarial Networks.** *arXiv:2106.12423*, <https://doi.org/10.48550/arXiv.2106.12423> (2021).

Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, Marx A, Boor P, Tacke F, Neumann UP, Grabsch HI, Yoshikawa T, Brenner H, Chang-Claude J, Hoffmeister M, Trautwein C, Luedde T. **Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer.** *Nat Med* 25, 1054–1056. <https://doi.org/10.1038/s41591-019-0462-y> (2019)

Kristiansen G, Delahunt B, Srigley JR, Lüders C, Lunkenheimer J-M, Gevensleben H, Thiesler T, Montironi R, Egevad L. **Vancouver-Klassifikation von Nierentumoren: Empfehlungen der Konsenskonferenz der Internationalen Gesellschaft für Uro-pathologie (ISUP) 2012.** *Die Pathologie, Ausgabe 3/2015* (2015).

Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. **Radiomics: The bridge between medical imaging and personalized medicine.** *Nat. Rev. Clin. Oncol.* 2017, 14, 729–762 (2017).

Lasseigne BN, Brooks JD. **The Role of DNA Methylation in Renal Cell Carcinoma.** *Mol Diagn Ther.* 2018 Aug;22(4):431-442. doi: 10.1007/s40291-018-0337-9. PMID: 29777398; PMCID: PMC6088752 (2018).

Li Z, Du G, Zhao R, Yang W, Li C, Huang J, Wen Z, Li H, Zhang B. Identification and validation of a hypoxia-related prognostic signature in clear cell renal cell carcinoma patients. *Medicine (Baltimore)*. 2021 Oct 1;100(39):e27374. doi: 10.1097/MD.00000000000027374. PMID: 34596153; PMCID: PMC8483867 (2021).

Long Q, Huang C, Huang J, Meng Q, Cheng Y, Li Y, He L, Chen M, Zhang C, Wang X, Zhu W, Peng J, Shi D, Zheng F, Dong P, Deng W. Prognostic value of JAK3 promoter methylation and mRNA expression in clear cell renal cell carcinoma. *Journal of Advanced Research*, Vol. 40, 2022, p.153-166, <https://doi.org/10.1016/j.jare.2021.11.016>. (2022).

McDermott DF, Choueiri TK, Motzer RJ, Aren OR, George S, Powles T, Donskov F, Harrison MR, Cid JRR, Ishii Y, McHenry MB, Mekan SF, Rini BI, Tannir NM. CheckMate 214 post-hoc analyses of nivolumab plus ipilimumab or sunitinib in IMDC intermediate/poor-risk patients with previously untreated advanced renal cell carcinoma with sarcomatoid features. *Journal of Clinical Oncology* 37, 4513–4513 (2019).

Moch H. Cystic renal tumors: new entities and novel concepts. *Adv Anat Pathol*. 2010 May;17(3):209-14. doi: 10.1097/PAP.0b013e3181d98c9d. PMID: 20418675 (2010).

Montani M, Heinimann K, von Teichman A, Rudolph T, Perren A, Moch H. VHL-gene deletion in single renal tubular epithelial cells and renal tubular cysts: further evidence for a cyst-dependent progression pathway of clear cell renal carcinoma in von Hippel-Lindau disease. *Am J Surg Pathol*. 2010 Jun;34(6):806-15. doi: 10.1097/PAS.0b013e3181ddf54d. PMID: 20431476 (2010).

Mühlbauer J, Egen L, Kowalewski KF, Grilli M, Walach MT, Westhoff N, Nuhn P, Laqua FC, Baessler B, Kriegmair MC. Radiomics in Renal Cell Carcinoma-A Systematic Review and Meta-Analysis. *Cancers (Basel)*. 2021 Mar 17;13(6):1348. doi: 10.3390/cancers13061348. PMID: 33802699; PMCID: PMC8002585 (2021).

Nilsson H, Lindgren D, Axelson H, Brueffer C, Saal LH, Lundgren J, Johansson ME. Features of increased malignancy in eosinophilic clear cell renal cell carcinoma. *J Pathol*. 2020 Dec;252(4):384-397. doi: 10.1002/path.5532. Epub 2020 Sep 24. PMID: 32815150; PMCID: PMC7756750 (2020).

Ohe C, Yoshida T, Amin MB, Uno R, Atsumi N, Yasukochi Y, Ikeda J, Nakamoto T, Noda Y, Kinoshita H, Tsuta K, Higasa K. Deep learning-based predictions of clear and eosinophilic phenotypes in clear cell renal cell carcinoma. *Human Pathology*, Volume 131, 2023, Pages 68-78, ISSN 0046-8177, <https://doi.org/10.1016/j.humpath.2022.11.004> (2023).

Oord AV, Vinyals O, Kavukcuoglu K. Neural Discrete Representation Learning. [Internet]. *arXiv[cs.LG]*. 2017. Available from: <http://arxiv.org/abs/1711.00937> (2017)

Pang S, Sun Y, Wu L, Yang L, Zhao Y-L, Wang Z, Li Y. Reconstruction of kidney renal clear cell carcinoma evolution across pathological stages. *Sci Rep* 8, 3339. <https://doi.org/10.1038/s41598-018-20321-4> (2018).

Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghouschi S, Anari S, Naseri M, Bendeche M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Sci Rep*. 2021 May 25;11(1):10930. doi: 10.1038/s41598-021-90428-8. PMID: 34035406; PMCID: PMC8149837 (2021).

Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, Bowlby R, Gibb EA, Akbani R, Beroukhim R, Bottaro DP, Choueiri TK, Gibbs RA, Godwin AK, Haake S, Hakimi AA, Henske EP, Hsieh JJ, Ho TH, Kanchi RS, Krishnan B, Kwiatkowski DJ, Lui W, Merino MJ, Mills GB, Myers J, Nickerson ML, Reuter VE, Schmidt LS, Shelley CS, Shen H, Shuch B, Signoretti S, Srinivasan R, Tamboli P, Thomas G, Vincent BG, Vocke CD, Wheeler DA, Yang L, Kim WY, Robertson AG, Cancer Genome Atlas Research Network, Spellman PT, Rathmell WK, Linehan WM. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep*. 2018 Apr 3;23(1):313-326.e5. doi: 10.1016/j.celrep.2018.03.075. Erratum in: *Cell Rep*. 2018 Jun 19;23(12):3698. PMID: 29617669; PMCID: PMC6075733 (2018).

Rini BI, Motzer RJ, Powles T, McDermott DF, Escudier B, Donskov F, Hawkins R, Bracarda S, Bedke J, De Giorgi U, Porta C, Ravaud A, Parnis F, Grande E, Zhang W, Huseni M, Carroll S, Sufan R, Schiff C, Atkins MB. Atezolizumab plus Bevacizumab Versus Sunitinib for Patients with Untreated Metastatic Renal Cell Carcinoma and Sarcomatoid Features: A Prespecified Subgroup Analysis of the IMmotion151 Clinical

Trial. *Eur Urol.* 2021 May;79(5):659-662. doi: 10.1016/j.eururo.2020.06.021. Epub 2020 Jul 9. PMID: 32654802; PMCID: PMC9376009 (2020).

Saldanha OL, Muti HS, Grabsch HI, Langer R, Dislich B, Kohlruss M, Keller G, van Treeck M, Hewitt KJ, Kolbinger FR, Veldhuizen GP, Boor P, Foersch S, Truhn D, Kather JN. **Direct prediction of genetic aberrations from pathology images in gastric cancer with swarm learning.** *Gastric Cancer* 26, 264–274 <https://doi.org/10.1007/s10120-022-01347-0> (2023).

Saliby RM, Saad E, Kashima S, Schoenfeld DA, Braun DA. Update on Biomarkers in Renal Cell Carcinoma. *Am Soc Clin Oncol Educ Book* 44, e430734(2024). DOI:10.1200/EDBK_430734

Schulz S, Woerl A-C, Jungmann F, Glasner C, Stenzel P, Strobl S, Fernandez A, Wagner D-C, Haferkamp A, Mildenberger P, Roth W and Foersch S. **Multimodal Deep Learning for Prognosis Prediction in Renal Cancer.** *Front. Oncol.* 11:788740. doi:10.3389/fonc.2021.788740 (2021).

Tomczak K, Czerwińska P, Wiznerowicz M. **The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge.** *Contemp Oncol (Pozn).* 2015;19(1A):A68-77. doi: 10.5114/wo.2014.47136. PMID: 25691825; PMCID: PMC4322527 (2015).

Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N. **Structure-Preserved Color Normalization for Histological Images.** *Proc – Int Symp BioMed Imaging* 35(8):1962–71. doi: 10.1109/ISBI.2015.7164042 (2015).

Vale-Silva LA, Rohr K. **Long-term cancer survival prediction using multimodal deep learning.** *Sci Rep* 11, 13505 (2021). <https://doi.org/10.1038/s41598-021-92799-4> (2021).

Vasudev NS, Wilson M, Stewart GD, Adeyolu A, Cartledge J, Kimuli M, Datta S, Hanbury D, Hrouda D, Oades G, Patel P, Soomro N, Sullivan M, Webster J, Selby PJ, Banks RE. **Challenges of early renal cancer detection: symptom patterns and incidental diagnosis rate in a multicentre prospective UK cohort of patients presenting with suspected renal cancer.** *BMJ Open* 2020;10:e035938. doi:10.1136/bmjopen-2019-035938 (2020).

Vuong L, Kotecha RR, Voss MH, Hakimi AA. **Tumor Microenvironment Dynamics in Clear-Cell Renal Cell Carcinoma.** *Cancer Discov.* 2019 Oct;9(10):1349-1357. doi: 10.1158/2159-8290.CD-19-0499. Epub 2019 Sep 16. PMID: 31527133; PMCID: PMC6774890 (2019).

Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, Sarveswara R, Händler K, Pickkers P, Aziz NA, Ktena S, Tran F, Bitzer M, Ossowski S, Casadei N, Herr C, Petersheim D, Behrends U, Kern F, Fehlmann T, Schommers P, Lehmann C, Augustin M, Rybniker J, COVID-19 Aachen Study (COVAS), DeCOI, Schultze JL. **Swarm Learning for decentralized and confidential clinical machine learning.** *Nature* 594, 265–270. <https://doi.org/10.1038/s41586-021-03583-3> (2021).

Zhang L, Zha Z, Qu W, Zhao H, Yuan J, Feng Y, Wu B. **Tumor necrosis as a prognostic variable for the clinical outcome in patients with renal cell carcinoma: a systematic review and meta-analysis.** *BMC Cancer* 18, 870. <https://doi.org/10.1186/s12885-018-4773-z> (2018).

Zhong W, Zhong H, Zhang F, Huang C, Lin Y, Huang J. **Characterization of Hypoxia-Related Molecular Subtypes in Clear Cell Renal Cell Carcinoma to Aid Immunotherapy and Targeted Therapy via Multi-Omics Analysis.** *Frontiers in Molecular Biosciences* 8, 2021, doi=10.3389/fmolb.2021.684050 (2021).

Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. **Learning Deep Features for Discriminative Localization.** 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2921-2929, doi: 10.1109/CVPR.2016.319 (2016).

5 Anhang: Publizierte Originalarbeit



Multimodal Deep Learning for Prognosis Prediction in Renal Cancer

Stefan Schulz¹, Ann-Christin Woerl^{1,2}, Florian Jungmann³, Christina Glasner¹, Philipp Stenzel¹, Stephanie Strobl¹, Aurélie Fernandez¹, Daniel-Christoph Wagner¹, Axel Haferkamp⁴, Peter Mildenberger³, Wilfried Roth¹ and Sebastian Foersch^{1*}

¹ Institute of Pathology, University Medical Center Mainz, Mainz, Germany, ² Institute of Computer Science, Johannes Gutenberg University Mainz, Mainz, Germany, ³ Department of Diagnostic and Interventional Radiology, University Medical Center Mainz, Mainz, Germany, ⁴ Department of Urology and Pediatric Urology, University Medical Center Mainz, Mainz, Germany

OPEN ACCESS

Edited by:

Viktor Grünwald,
Universitätsklinikum Essen, Germany

Reviewed by:

Hendrik Eggers,
Hannover Medical School, Germany
Henning Reis,
Goethe University Frankfurt, Germany

*Correspondence:

Sebastian Foersch
sebastian.foersch@
unimedizin-mainz.de

Specialty section:

This article was submitted to
Genitourinary Oncology,
a section of the journal
Frontiers in Oncology

Received: 03 October 2021

Accepted: 02 November 2021

Published: 24 November 2021

Citation:

Schulz S, Woerl A-C, Jungmann F,
Glasner C, Stenzel P, Strobl S,
Fernandez A, Wagner D-C,
Haferkamp A, Mildenberger P, Roth W
and Foersch S (2021) Multimodal
Deep Learning for Prognosis
Prediction in Renal Cancer.
Front. Oncol. 11:788740.
doi: 10.3389/fonc.2021.788740

Background: Clear-cell renal cell carcinoma (ccRCC) is common and associated with substantial mortality. TNM stage and histopathological grading have been the sole determinants of a patient's prognosis for decades and there are few prognostic biomarkers used in clinical routine. Management of ccRCC involves multiple disciplines such as urology, radiology, oncology, and pathology and each of these specialties generates highly complex medical data. Here, artificial intelligence (AI) could prove extremely powerful to extract meaningful information to benefit patients.

Objective: In the study, we developed and evaluated a multimodal deep learning model (MMDLM) for prognosis prediction in ccRCC.

Design, Setting, and Participants: Two mixed cohorts of non-metastatic and metastatic ccRCC patients were used: (1) The Cancer Genome Atlas cohort including 230 patients and (2) the Mainz cohort including 18 patients with ccRCC. For each of these patients, we trained the MMDLM on multiscale histopathological images, CT/MRI scans, and genomic data from whole exome sequencing.

Outcome Measurements and Statistical Analysis: Outcome measurements included Harrell's concordance index (C-index) and also various performance parameters for predicting the 5-year survival status (5YSS). Different visualization techniques were used to make our model more transparent.

Results: The MMDLM showed great performance in predicting the prognosis of ccRCC patients with a mean C-index of 0.7791 and a mean accuracy of 83.43%. Training on a combination of data from different sources yielded significantly better results compared to when only one source was used. Furthermore, the MMDLM's prediction was an independent prognostic factor outperforming other clinical parameters.

Interpretation: Multimodal deep learning can contribute to prognosis prediction in ccRCC and potentially help to improve the clinical management of this disease.

Patient Summary: An AI-based computer program can analyze various medical data (microscopic images, CT/MRI scans, and genomic data) simultaneously and thereby predict the survival time of patients with renal cancer.

Keywords: artificial intelligence, deep learning, pathology, prognosis prediction, radiology, renal cancer

INTRODUCTION

Clear-cell renal cell carcinoma (ccRCC) is the most common type of kidney cancer and more than 175,000 patients die from this entity each year (1). In contrast to other tumor types, there is no clearly defined set of biomarkers used in clinical routine. This might be partly because ccRCC development seems to be driven by a multitude of interacting metabolic pathways and regulated by complex epigenetic programs (2). Clinical management of ccRCC usually involves various specialties including urology, radiology, oncology, pathology, and others. This results in a vast amount of medical data on each patient, such as CT/MRI scans, histopathological images, and other clinical information. There are several clinical tools for prognosis prediction in ccRCC, such as the UCLA integrated Staging System (UISS) (3) or the risk model of the International Metastatic RCC Database Consortium (IMDC) (4). Heng et al. for example developed a score, which consists of various clinical parameters such as Karnofsky performance status, hemoglobin, corrected calcium, and others. With this strategy, they were able to achieve an overall C-index of 0.73 in the prognosis prediction of 645 metastatic RCC (4). But while prognostic clinical nomograms might be helpful, they can be cumbersome to use and often only incorporate a selection of the available information—both of which potentially limit their performance. Here, artificial intelligence (AI) and machine learning (ML) could prove extremely helpful to utilize these highly complex data to predict clinically relevant outcomes such as survival or therapy response.

AI and ML are increasingly being applied to various medical problems achieving highly promising results in ophthalmology (5), radiology (6), cardiology (7), and others—even surpassing human level performance in some cases (8). For pathological tasks for example, we were able to predict the molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides alone using deep learning (DL) (9). We also used a similar approach for prognosis prediction in soft tissue sarcoma (STS) (10). But while it is technically feasible, there are very few studies so far evaluating the use of multimodal input for training of AI and DL models (11). Thus, we developed a comprehensive DL pipeline and utilized multiscale conventional histopathological images together with CT/MRI images and genomic data to predict survival in patients with ccRCC.

MATERIALS AND METHODS

Patient Cohorts

Two cohorts were utilized in which patients with ccRCC were included. The first cohort served as the basis for training of the

neural network and validation to determine performance metrics. It consisted of all patients of the KIRC TCGA (Kidney renal clear cell carcinoma of the Cancer Genome Atlas) cohort for which the diagnostic H&E (hematoxylin & eosin) stained whole slide as well as radiological images were available. These were downloaded for 230 patients through the GDC portal (<https://portal.gdc.cancer.gov/>) as well as from the cancer imaging archive (<https://www.cancerimagingarchive.net/>). All initial pathology reports, clinicopathological and survival data (disease-specific survival, DSS) as well as the ten most frequent mutations/copy number alterations in our cohort were gathered from www.cbioportal.org. A comprehensive quality assessment excluded histopathological slides with large folds, no tumor tissue, and/or where the image was out of focus. Whenever possible computer tomography (CT) scans with nephrogenic or late systemic arterial phase were used. In a subset of patients, only magnet resonance imaging (MRI) scans were available. In this case T1-weighted sequences were used when possible. A second, mono-center cohort of 18 patients was generated as an additional external test set (the Mainz cohort). These patients were diagnosed between 2011 and 2015 at the University Medical Center Mainz. Retrospective use of these and other patients' data and material for research purposes was approved by the ethical committee of the medical association of the State of Rhineland-Palatinate [Ref. Nos. 837.360.16 (10679) and 837.031.15(9799)] and results were generated after 2-fold pseudonymization of the cohort. We settled on this relatively low number of patients to ensure high quality of the radiologic, pathologic, and clinical follow up data. Tumor staging, grading, and treatment for these patients was carried out according to the appropriate guidelines in place at that time (i.e., ISUP). All experiments were in accordance with the Declaration of Helsinki (**Supplementary Tables 1, 2**).

Scanning and Preprocessing

TCGA whole slide images (WSIs) were digitalized at various institutions participating in the TCGA consortium. Slides from the second cohort were scanned using a Hamamatsu Nanozoomer Series scanner (Hamamatsu Photonics, Hamamatsu, Japan) at 40-fold magnification. This translated to a resolution of 0.23 $\mu\text{m}/\text{pixel}$. Slides were thoroughly evaluated for routine diagnostics by a board-certified pathologist and annotated by the project team blinded to any of the target variables. Annotation describes the process in which a polygonal region of interest was drawn around the tumor area. Various tissue aspects were considered (i.e., necrosis, etc.). Image tiles (520 px^2) from two magnification levels (level 5 $\approx 10\times$ magnification and level 10 $\approx 5\times$ magnification) were then generated from these annotations. All tiles were normalized to

an external reference image of a different dataset using structure preserving color normalization (SPCN) as proposed by Vahadane et al. (12) and Anand et al. (13). CT/MRI scans were also gathered at the respective institutions participating in the TCGA consortium. CT scans from the second cohort were generated using 64- or 128-section CT systems (Philips, Eindhoven, Netherlands) and 1.5T MRI scanner (Siemens, Forchheim, Germany). Scans were thoroughly evaluated for routine diagnostics by a board-certified radiologist and annotated by the project team blinded to any of the target variables. For each 3D volume, three images were extracted showing the maximum tumor diameter (one transversal plane, one sagittal plane and one coronal plane). Examples for annotation, tiling, normalization, and augmentation can be found in **Supplementary Figures 1A–C**.

Novel Deep Learning Pipeline

A new, comprehensive, multimodal deep learning model (MMDLM) was developed consisting of one individual 18-layer residual neural network (ResNet) per image modality (resulting in up to three ResNets) and a dense layer for genomic data. This particular architecture was chosen to compromise between model depth and computational time. After this, the network outputs were concatenated by an attention layer, which weights every output due to its importance for the given task and passed through a final fully connected network, followed by either C-index calculation (see below) or binary classification (5-year disease-specific survival status (5YSS)). The 5YSS includes all patients who either lived longer than 60 months or who passed away within five years after diagnosis. All patients for whom the follow-up time was shorter than five years were not included in this analysis, as one cannot be certain, whether these patients would have survived five years or longer. Unimodal training was performed by muting all other inputs and initializing ResNet weights on pretrained image net weights. Multimodal training was then carried out using the pretrained weights of the unimodal training. A total of 200–400 epochs were trained, and the best model was chosen when training and validation curves stopped converging. Standard Cox loss was used as loss function and cross-entropy loss was used as loss function for binary classification. Cox loss is defined as

$$l_{cox}(\theta) = -\sum_{i: E_i=1} \left(\hat{h}_{\theta}(x_i) - \log \sum_{j: T_j > T_i} e^{\hat{h}_{\theta}(x_j)} \right)$$

Where T_i represents the survival time, E_i the censored event status, x_i the data for each patient, and \hat{h}_{θ} the neural network model trained (14). Stochastic gradient descent was used as optimizer, learning rate was set at 0.004, momentum was set at 0.9, batch size was 32. Training was performed with a customized data loader which generated random combinations of one histopathologic image at level 5, one at level 10 and one radiologic image for each patient. Genomic data could optionally be included, but we limited the number to the 10 most frequent mutations, not to make the model overly complex. Validation was performed on a patient level using the Cartesian product of primary fixed image combinations to make results more comparable. Classification markup was performed using our previously reported sliding-window approach after training

of an unmodified 18-layer ResNet (9). Class activation maps (CAMs) were established as recently described (15).

Statistical Analysis

Training and validation on the TCGA cohort were performed using full k-fold cross validation (CV) on a patient level (6-fold CV for C-index prediction and 12-fold CV for binary classification). Metrics included C-index, recall (sensitivity), true negative rate (specificity), precision, area under the curve (AUC) of the receiver operating characteristic (ROC) and also the precision recall curve (PRC). Concordance index (C-index) was calculated as implemented by the lifelines package (16). In short, it is a measurement of the ability of a model to rank each patient according to their actual survival times based on individual risk scores. We used the C-index implementation by Davidson-Pilon et al. (16) as the number of censored events was not unreasonably high and as it was the easiest to integrate into our setup. For the 5YSS all patients surviving longer than 5 years were compared to all patients who died within the first 5 years. Patients lost to follow up within the first five years were not included in these analyses. The mean AUC of ROC either of multiple classes or as a summary of cross validation for each individual class was calculated using micro- and macro-averaging (17). For each analysis, the values' distribution was tested. Paired t-test was used when two individual groups with normal distribution of paired experiments were analyzed. One-sample t-test was used to compare column means to a single value. Repeated measures (RM) one-way ANOVA with *post-hoc* Tukey HSD to correct for multiple comparisons was used when more than two groups with normal distribution were compared. Log-rank test was used for comparison of two or more survival curves. Univariable and multivariable Cox regression was used for prognosis analyses after checking proportionality using scaled Schoenfeld residuals. If not indicated otherwise, \pm standard deviation (SD) is given. Differences in the compared groups were considered statistically significant when P values were smaller than 0.05 ($p \geq 0.05$: ns, $p = 0.01-0.05$ *, $p = 0.001-0.01$ ***, $p = 0.0001-0.001$ ****, $p < 0.0001$:****). For annotations and image preprocessing of the histopathologic slides, QuPath open source software (18) was used. For annotation and image processing of the radiologic volumes, Mango (19) and 3D Slicer (20) were used. All deep learning experiments were done in Python using PyTorch/fast.ai or TensorFlow/Keras. Statistical analysis was done using Graph Pad Prism or R. Some images were created with BioRender.com. Our algorithms were developed utilizing open access material and tutorials, such as PyImageSearch by Adrian Rosebrock, "Practical Deep Learning for Coders" by Jeremy Howard, and others. Code samples, etc. might be provided within collaboration with the project team. Please contact the corresponding author.

RESULTS

Clinicopathological features of the TCGA cohort can be found in **Figure 1A**. A total of 58,829 tiles at level 5 and 17,514 tiles at level

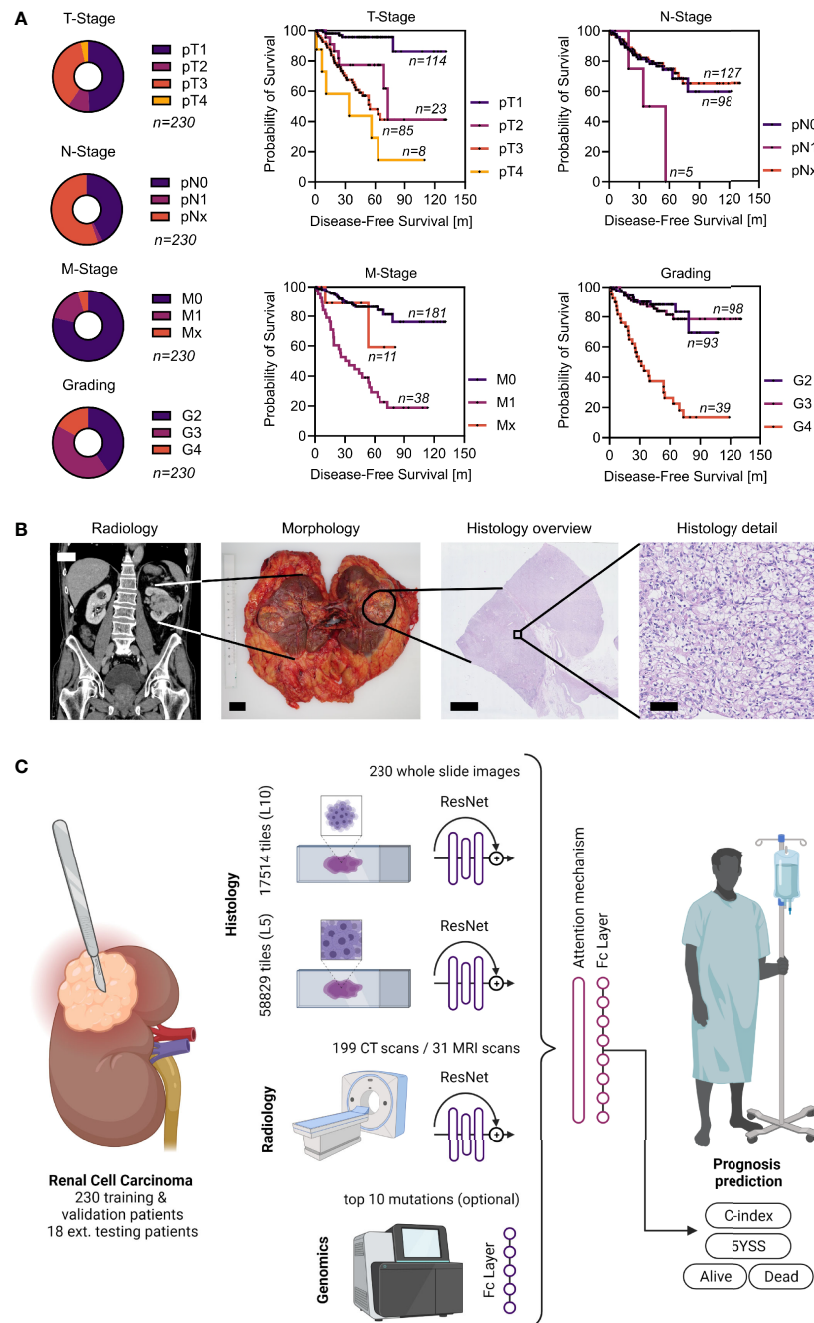


FIGURE 1 | Patient cohort, clinical example, and overview of the MMDLM. **(A)** Characteristics of the TCGA cohort. **(B)** Clinical example of a typical ccRCC case. CT (scalebar 5 cm), macroscopic (scalebar 2 cm), as well as histologic tumor appearance (scalebars 5 mm and 100 μ m) are displayed. **(C)** Schematic overview of the model. Created with BioRender.com.

10 were generated from 230 whole slide images for the training and validation experiments. Approximately 199 CT scans as well as 31 MRI scans from the same cohort were used to generate a total of 690 coronal, sagittal and transversal images. A typical example of a ccRCC case and a scheme of the MMDLM is displayed in **Figures 1B, C**.

First, we wanted to establish a baseline of the prognosis prediction capabilities of each imaging modality alone. To this end we calculated the C-index, which is a measurement of the ability of a model to rank each patient according to their actual survival times based on individual risk scores. Using unimodal training on radiological data yielded a mean C-index

of 0.7074 ± 0.0474 with a maximum of 0.7590. Training only on histopathological image tiles our model achieved a mean C-index of 0.7169 ± 0.0296 with a maximum of 0.7638 (level 10) and a mean C-index of 0.7424 ± 0.0339 with a maximum of 0.7821 (level 5), respectively. Next, we wanted to investigate, whether the combination of different imaging modalities would improve prognostication in ccRCC. When combining conventional histopathological input with CT and MRI images, the mean C-index increased to 0.7791 ± 0.0278 with a maximum of 0.8123. There was a significant difference when compared to C-index of training only on radiologic images (p-value = 0.0207) and histopathologic tiles (p-value = 0.0140) (Figure 2A). Next, we wanted to investigate how the uni- and multimodal deep learning models performed, when compared to known prognostic factors in renal cell carcinoma. Cox proportional hazard (CPH) models were used to calculate the C-indices for histopathological grading (0.7010), T-Stage (0.7470), N-Stage (0.5140), and M-Stage (0.6850). Strikingly, only the MMDLM was significantly better than all independent prognostic factors (Figure 2B).

Since the C-index cannot be applied to an individual patient and thus might prove difficult to be translated into clinical decision making, we investigated the possibility to predict the 5-year survival status (5YSS) using a MMDLM and binary classification. A total of

113 patients could be included in these analyses. Here accuracy reached $83.43\% \pm 11.62\%$ with a maximum of 100% upon 12-fold cross validation. This was higher, when compared to unimodal approaches, however this did not reach statistical significance. AUC of the ROC was 0.916 ± 0.105 with a maximum of 1.0. AUC of the PR curve was 0.944 ± 0.075 with a maximum of 1.0 (Figure 2C). Dividing the cohort according to the MMDLM's prediction ("Alive" vs. "Dead") into low- and high-risk patients showed a highly significant difference in the survival curves (Figure 2D). This was also true when only non-metastasized (M0) or metastasized (M+) patients were evaluated (Supplementary Figure 2). To compare the MMDLM's prediction with the known risk factors described above, we performed multivariable regression analyses. Here only T-Stage and MMDLM's prediction showed to be independent, significant prognostic factors with the MMDLM displaying the highest hazards ratio of almost 4 (Figure 2E).

To investigate whether the addition of genomic data could further improve our image-based prognosis prediction, we compared the performance of the MMDLM with and without training on the top ten mutations/copy number alterations (CNA) found in our cohort (Figure 3A). Interestingly, there was no improvement by adding this type of information to the training process. Looking at all alterations together or each alteration separately, none was able to show a statistically

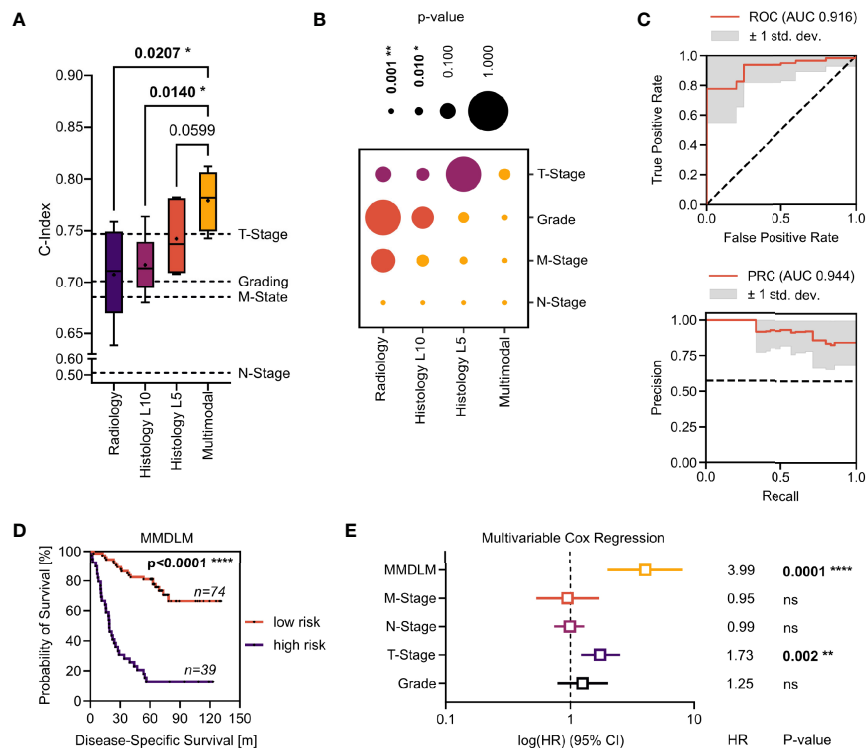
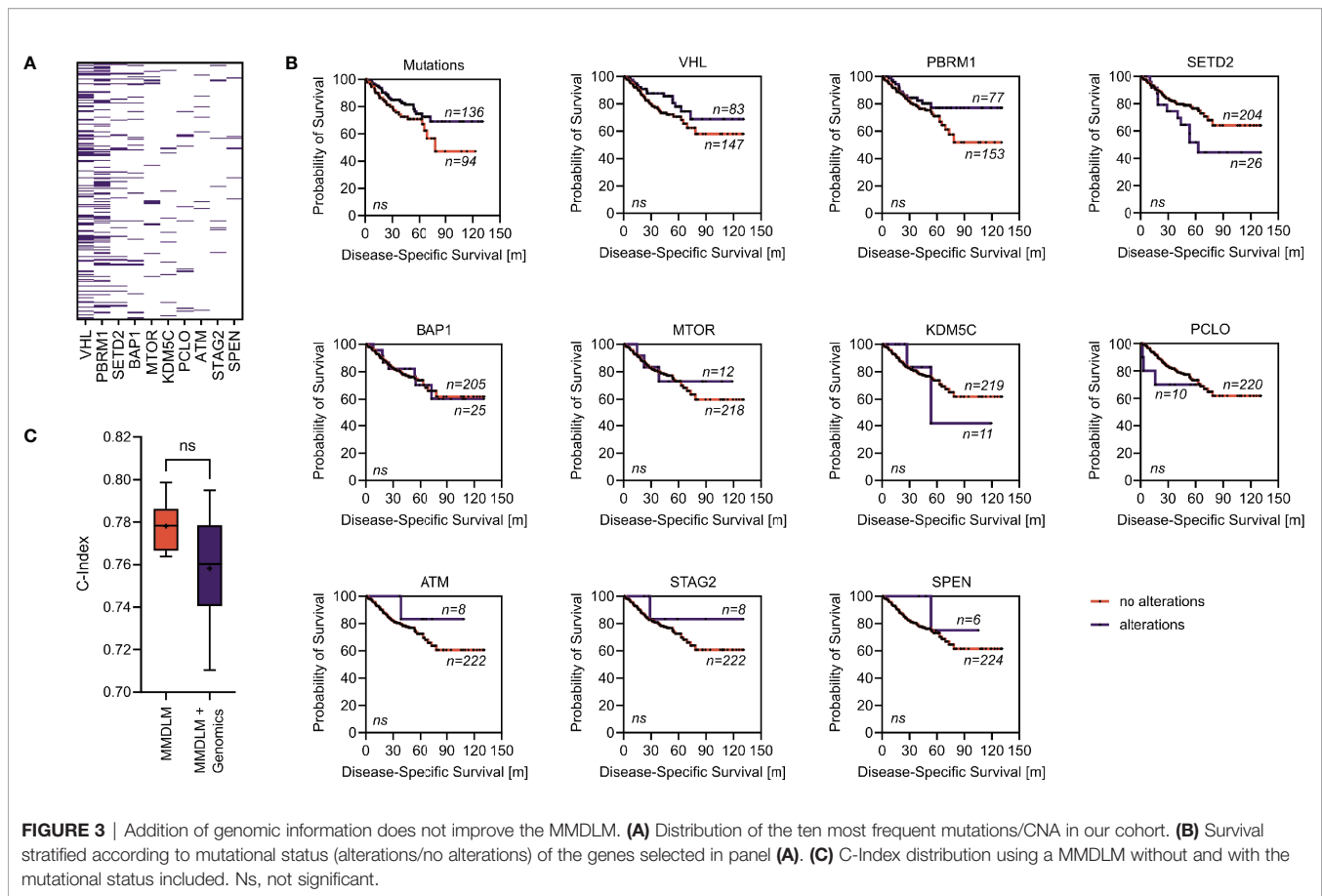


FIGURE 2 | Evaluation of the MMDLM for prognosis prediction in ccRCC. **(A)** C-index distribution of 6-fold cross validation. Dotted lines represent the C-index of the respective clinical attribute (Grading, T-Stage, N-Stage, M-Stage) of the whole cohort. RM one-way ANOVA with *post-hoc* Tukey HSD to correct for multiple comparisons was used to compare the groups. **(B)** P-value matrix of one-sample t test of each modality vs. each risk factor (yellow: significantly higher, orange: higher, purple: lower). **(C)** Mean ROC (top) and PR curve (bottom) of 12-fold cross validation. **(D)** Kaplan-Meier-Curve after stratification according to 5YSS by the MMDLM. **(E)** Forrest plot of multivariable Cox regression. HR, hazard ratio; CI, Confidence interval; Ns, not significant. *p = 0.01–0.05, **p = 0.001–0.01, ****p < 0.0001



significant difference in survival of patients with ccRCC (Figures 3B, C).

Since we only trained and evaluated on the TCGA dataset thus far, we wanted to investigate how the MMDLM would perform on an additional external test set. This consisted of 18 patients representing 9.3% of the training set for C-index calculation and 17.6% for binary classification. Mean C-index reached 0.799 ± 0.060 with a maximum of 0.8662. Accuracy averaged at $79.17\% \pm 9.8\%$ with a maximum of 94.44%. AUC of the ROC was 0.905 ± 0.073 with a maximum of 1.0. AUC of the PR curve was 0.956 ± 0.036 with a maximum of 1.0. All performance measures were not significantly different from those achieved during cross validation (CV) (Supplementary Figure 3).

Lastly, we aimed at increasing the transparency of our model by visualizing the image features for each modality that were most relevant to the model's prediction. We used a sliding window approach to visualize unimodal classification WSIs (Figure 4A). We established class activation maps (CAMs) using the CV fold with the highest C-index prediction, consisting of 17,550 image combinations. Investigating these image combinations of all patients of this fold, a first descriptive screening analysis of representative CAMs revealed histopathologic (such as tumor vasculature, hemorrhage, and necrosis) and radiologic (such as tumor volume) features which were most important to the model to make its prediction (Figure 4B).

DISCUSSION

Diagnosis and treatment of ccRCC remains a clinical challenge—especially in metastasized cases. For both non-metastasized and metastasized patients, prognostic tools exist such as UISS and IMDC, but there is still room for improvement. In our study, we propose a MMLDM, which could be a valuable alternative and/or addition to existing tools in both M0/M+ patients. While methods of AI are increasingly being used in various medical domains, their combination across different modalities has only rarely been explored (21). This is particularly surprising as such combination efforts are already being developed and deployed in non-medical fields such as autonomous driving and others. Furthermore, the few multimodal or “fusion” approaches applied to medical problems consistently showed a boost in accuracy of up to 27.7% when compared over single modality models for the same task (22). However, most of this work is limited to the integration of low-level clinical features with one type of imaging data to make a certain diagnosis. Here we describe three major improvements over most previous studies. (I) A variety of comprehensive histopathologic and radiologic imaging techniques together with genetic information derived from whole exome sequencing were integrated in our model. This mirrors the clinical decision-making process (i.e., during interdisciplinary tumor boards) and was done to ensure, that as

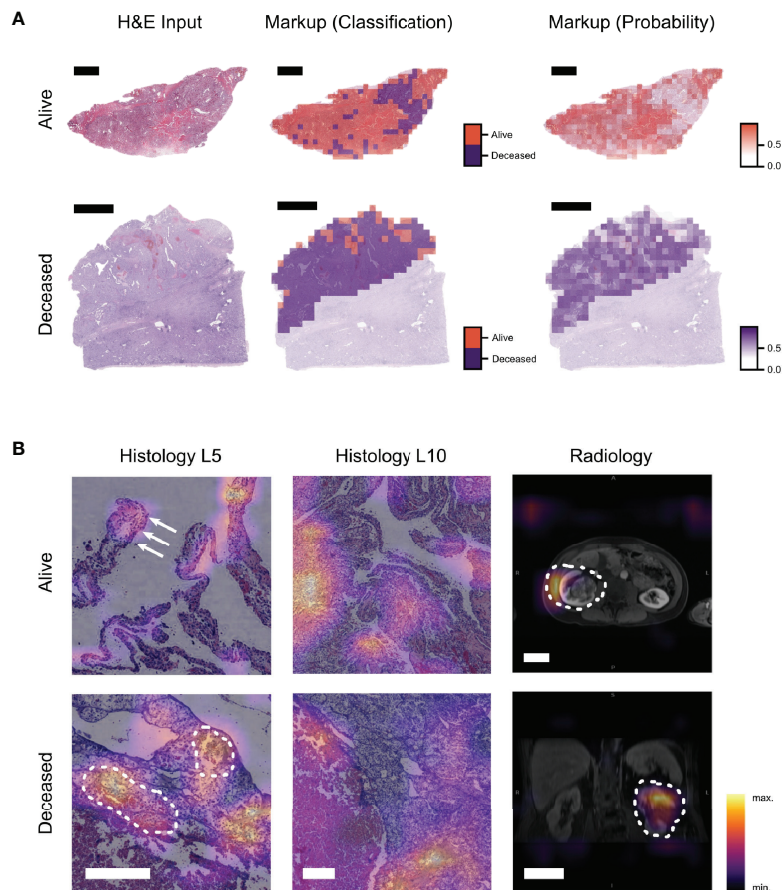


FIGURE 4 | Visualization techniques show image regions important for the prediction and their contribution to the MMDLM. **(A)** Example of a visualization approach to display the classification result of a unimodal histopathology model (ResNet18—Level 5). The input WSI as well as two different markup images are displayed. Markup all denotes the distinction between tiles classified as alive or deceased. Markup class denotes the prediction certainty within the majority class (scalebar top row: 4 mm, scalebar bottom row: 5 mm). **(B)** CAMs of the MMDLM are shown. Different features associated with low-risk (alive) and high-risk (deceased) are highlighted. In the low-risk example, clear cell morphology as well as papillary tumor appearance (arrows) can be observed. In the high-risk example, tumor vasculature and bleeding can be observed (dotted line) (scalebar histology: 250 μ m, scalebar radiology: 5 cm).

much of the relevant information was utilized as possible. (II) Target variable was not the diagnosis of a certain tumor entity but rather the prognosis of the patients. This is particularly relevant in renal cancer as there is an urgent need for reliable prognostic biomarkers in this entity. Our integrative approach could be used to distinguish between low- and high-risk patients, who would be more suitable for intensified treatment and/or surveillance. Interestingly, the addition of genomic data did not improve the image-based multimodal approach—highlighting the fact that mutations/CNA are of less prognostic value in our cohort. This might be because ccRCC is highly dependent on mutations that are very common in this tumor type. (III) We also used additional visualization techniques to highlight image features which were most relevant to our model.

Liu et al. used photographs together with clinical data to classify skin lesions and showed that the top-1-accuracy of their deep learning system was even slightly better than the one of trained dermatologists (23). However, to achieve this accuracy

the group had to use data from over 15,000 patients, which might not be easily accessible for every clinical question. Furthermore, while the authors show how training only on images decreases the model's performance it is unclear how the model would have performed on clinical data alone. By using CAMs, we were able to investigate image features associated with prognosis, although only in a descriptive fashion thus far. A recent publication by Ning et al. uses convolutional neural networks (CNNs) for feature extraction on radiologic and pathologic data, and combines these features with genomic data for prognosis prediction in ccRCC (24) with similar results. However, in the study by Ning et al. it remains unclear how and which image features were selected and how the model would perform on a true external test set. Of course, there are limitations to our approach as well. For example the comparison between other clinical tools, which include clinical data such as performance status, calcium levels, etc. are missing. So a head to head comparison with IMDC or UISS scores is necessary to

determine superiority of our MMDLM. Furthermore, the size of the external validation is rather small, and additional studies are needed to ensure generalizability of our approach.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of institutional requirements and general data protection regulation. The WSI for the TCGA Dataset are publicly available from the Genomic Data Commons Data Portal. Requests to access the datasets should be directed to sebastian.foersch@unimedizin-mainz.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethical committee of the medical association of the State of Rhineland-Palatinate [Ref. No. 837.360.16(10679) and 837.031.15(9799)]. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conception and design: SF, SSc, A-CW. Acquisition of data: SF, SSc, A-CW, FJ, CG, PS, SSt, AH, and PM. Analysis and interpretation of data: SF, SSc, A-CW, FJ, CG, PS, SSt, D-CW, PM, and WR. Drafting of the manuscript: SF and SSc. Critical revision of the manuscript for important intellectual content: A-CW, FJ, PS, AH, PM, and WR. Statistical analysis: SF, SSc, A-CW, and CG. Obtaining funding: SF. Administrative, technical, or material support: AF. Supervision: SF and WR. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Federal Ministry of Education and Research (16SV8167), the Stage-I-Program of the University

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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 (6):394–424. doi: 10.3322/caac.21492
- Cancer Genome Atlas Research Network. Comprehensive Molecular Characterization of Clear Cell Renal Cell Carcinoma. *Nature* (2013) 499 (7456):43–9. doi: 10.1038/nature12222
- Zisman A, Pantuck AJ, Wieder J, Chao DH, Dorey F, Said JW, et al. Risk Group Assessment and Clinical Outcome Algorithm to Predict the Natural History of Patients With Surgically Resected Renal Cell Carcinoma. *J Clin Oncol* (2002) 20(23):4559–66. doi: 10.1200/JCO.2002.05.111
- Heng DY, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C, et al. Prognostic Factors for Overall Survival in Patients With Metastatic Renal Cell Carcinoma Treated With Vascular Endothelial Growth Factor-Targeted Agents: Results From a Large, Multicenter Study. *J Clin Oncol* (2009) 27 (34):5794–9. doi: 10.1200/JCO.2008.21.4809
- De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically Applicable Deep Learning for Diagnosis and Referral in

Medical Center Mainz, the Mainz Research School of Translational Biomedicine (TransMed), and the Manfred-Stolte-Foundation.

ACKNOWLEDGMENTS

Aspects of this work are part of the MD thesis of SSc.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.788740/full#supplementary-material>

Supplementary Figure 1 | Data preprocessing. (A) Preprocessing of the input WSI (top) included annotation (middle) and tiling into level 5 tiles (bottom)(scalebar 2 mm). (B) Input tiles (top) were normalized (middle) to a reference image not from the cohorts investigated (scalebar 250 μ m) and augmented (bottom) during training. (C) Preprocessing of the radiologic data included export of a coronal, transversal, and sagittal image with the largest tumor diameter ("pseudovolume"). Icon modified from CFCF, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=33285529>.

Supplementary Figure 2 | Survival of MMDLM low- and high-risk cases according to M0/M+. (A) Kaplan-Meier-Curve after stratification according to 5YSS by the MMDLM for M0 patients. (B) Kaplan-Meier-Curve after stratification according to 5YSS by the MMDLM for M+ patients.

Supplementary Figure 3 | Evaluation of the MMDLM on an additional external test cohort. (A) Frequency distribution of the additional test cohort (the Mainz cohort) is similar to the TCGA cohort. (B) Mean C-index is not significantly different between the validation and the additional external test set. (C) Accuracy is not significantly different between the validation and the external test set. (D) ROC and PR curves of the external test set. Ns, not significant.

Supplementary Table 1 | TCGA cohort.

Supplementary Table 2 | Additional external test cohort (the Mainz cohort).

- Retinal Disease. *Nat Med* (2018) 24(9):1342–50. doi: 10.1038/s41591-018-0107-6
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-To-End Lung Cancer Screening With Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography. *Nat Med* (2019) 25(6):954–61. doi: 10.1038/s41591-019-0447-x
- Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic Diagnosis of the 12-Lead ECG Using a Deep Neural Network. *Nat Commun* (2020) 11(1):1–9. doi: 10.1038/s41467-020-15432-4
- Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056
- Woerl A-C, Eckstein M, Geiger J, Wagner DC, Daher T, Stenzel P, et al. Deep Learning Predicts Molecular Subtype of Muscle-Invasive Bladder Cancer From Conventional Histopathological Slides. *Eur Urol* (2020) 78(2):256–64. doi: 10.1016/j.eururo.2020.04.023
- Foersch S, Eckstein M, Wagner D-C, Gach F, Woerl AC, Geiger J, et al. Deep Learning for Diagnosis and Survival Prediction in Soft Tissue Sarcoma. *Ann Oncol* (2021) 32(9):1178–87. doi: 10.1016/j.annonc.2021.06.007

11. Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. *IEEE Trans Med Imaging* (2020), 1–1. doi: 10.1109/TMI.2020.3021387
12. Vahadane A, Peng T, Albarqouni S, Albarqouni S, Wang L, Baust M, et al. Structure-Preserved Color Normalization for Histological Images. *Proc - Int Symp BioMed Imaging* (2015) 35(8):1962–71. doi: 10.1109/ISBI.2015.7164042
13. Anand D, Ramakrishnan G, Sethi A. Fast GPU-Enabled Color Normalization for Digital Pathology. *Int Conf Syst Signals Image Process* (2019) 2019:219–24. doi: 10.1109/IWSSIP.2019.8787328
14. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network. *BMC Med Res Methodol* (2018) 18:1–12. doi: 10.1186/s12874-018-0482-1
15. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* (2016) 2016:2921–9. doi: 10.1109/CVPR.2016.319
16. Davidson-Pilon C. CamDavidsonPilon/lifelines: V0.25.8 (Version V0.25.8). *Zenodo* (2021). doi: 10.5281/zenodo.4457577
17. Sokolova M, Lapalme G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf Process Manag* (2009) 45(4):427–37. doi: 10.1016/j.ipm.2009.03.002
18. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci Rep* (2017) 7(1):16878. doi: 10.1038/s41598-017-17204-5
19. Lancaster JL, Cykowski MD, McKay DR, Kochunov PV, Fox PT, Rogers W, et al. Anatomical Global Spatial Normalization. *Neuroinformatics* (2010) 8(3):171–82. doi: 10.1007/s12021-010-9074-x
20. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001
21. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, et al. Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks. *Proc Natl Acad Sci USA* (2018) 115(13):E2970–9. doi: 10.1073/pnas.1717139115
22. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of Medical Imaging and Electronic Health Records Using Deep Learning: A Systematic Review and Implementation Guidelines. *NPJ Digit Med* (2020) 3(1):136. doi: 10.1038/s41746-020-00341-z
23. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A Deep Learning System for Differential Diagnosis of Skin Diseases. *Nat Med* (2020) 26(6):900–8. doi: 10.1038/s41591-020-0842-3
24. Ning Z, Pan W, Chen Y, Xiao Q, Zhang X, Luo J, et al. Integrative Analysis of Cross-Modal Features for the Prognosis Prediction of Clear Cell Renal Cell Carcinoma. *Bioinformatics* (2020) 36(9):2888–95. doi: 10.1093/bioinformatics/btaa056

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Schulz, Woerl, Jungmann, Glasner, Stenzel, Strobl, Fernandez, Wagner, Haferkamp, Mildenerberger, Roth and Foersch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

6 Danksagung

Hiermit möchte ich allen Koautoren und Koautorinnen für Ihre wertvollen Beiträge, Ratschläge und die gute Zusammenarbeit danken, welche die vorliegende Studie erst ermöglicht haben.

Herrn ... und Herrn ... möchte ich für die Bereitstellung der radiologischen Bilddaten danken.

Herrn ... und Herrn ... danke ich besonders für die ausgezeichnete Betreuung und Supervision während der Durchführung der Arbeit.

7 Tabellarischer Lebenslauf