

ZUSAMMENFASSUNG

Patrick Schelb
Dr. med.

Entwicklung eines Deep-Learning-basierten Verfahrens zur automatisierten Prostatakarzinomdiagnostik mittels biparametrischer Magnetresonanztomographie

Fach/Einrichtung: Deutsches Krebsforschungszentrum (DKFZ)
Doktorvater: Prof. Dr. med. David Bonekamp

Das Prostatakarzinom stellt in der westlichen Welt das häufigste bösartige Karzinom und eine der häufigsten tumorbedingten Todesursachen von Personen mit einer Prostata dar. Die MRT stellt ein vielversprechendes Verfahren dar, um die Prostatakarzinomdiagnostik zu verbessern und invasive Diagnostik zu reduzieren.

Es wurde die Fragestellung untersucht, ob ein Deep-Learning-basierter Segmentierungsansatz aufgrund T2-gewichteter und diffusionsgewichteter MRT- Bilder eine Prostatakarzinomdiagnose stellen kann, die vergleichbar zur klinischen MRT-Interpretation ist.

Hierfür wurde retrospektiv eine konsekutive Kohorte aus 312 Personen (mittleres Alter, 64 Jahre, Interquartilsabstand 58–71 Jahre) eingeschlossen, die zwischen 2015 und 2016 eine MRT-Untersuchung an demselben 3.0-Tesla-Gerät bekommen hatten und die klinische Indikation für eine Prostatabiopsie besaßen. Läsionen, die während der klinischen Befundung durch das *Prostate Imaging Reporting and Data System* auffielen, wurden manuell segmentiert. Alle eingeschlossenen Personen erhielten eine kombinierte gezielte und systematische MRT-gesteuerte Ultraschall-Fusionsbiopsie. Das klinisch signifikante Prostatakarzinom wurde definiert als *International Society of Urological Pathology Gleason Grade Group* ≥ 2 . Ein U-Net-basierter Deep-Learning- Ansatz wurde mithilfe einer mehrfach wiederholten Kreuzvalidierung auf der Trainingskohorte (80 % der Gesamtkohorte) entwickelt und validiert. Im Anschluss wurde die Wertigkeit des U-Nets auf einer ungesesehenen Testkohorte (20 % der Gesamtkohorte) getestet. Untersucht wurde sowohl die Leistung auf Personenlevel als auch auf Sextantenlevel und Läsionslevel. Die Arbeitspunkte des U-Nets wurden auf der Sextanten-basierten Analyse kalibriert, um möglichst nahe die Leistung der klinischen Befundung durch das *Prostate Imaging Reporting and Data System* widerzuspiegeln. Für den endgültigen Vergleich zwischen U-Net und radiologischer Befundung wurden Sensitivität, Spezifität, Positiv Prädiktiver Wert, Negativ Prädiktiver Wert und der Dice-Wert miteinander verglichen. Ein Signifikanzniveau von 0,05 wurde als statistisch signifikant angesehen.

Auf der Testkohorte hatte ein Prostate-Imaging-Reporting-and-Data-System-Wert ≥ 3 auf Personenbasis eine Sensitivität von 96 % (25 von 26) und eine Spezifität von 22 % (8 von 36). Prostate-Imaging-Reporting-and-Data-System-Wert ≥ 4 hatte eine Sensitivität von 88 % (23 von 26) und eine Spezifität von 50 % (18 von 36). U-Net hatte am Arbeitspunkt $\geq 0,22$ eine Sensitivität von 96 % (25 von 26) und eine Spezifität von 31 % (11 von 36), sowie am Arbeitspunkt $\geq 0,33$ eine Sensitivität von 92 % (24 von 26) und eine Spezifität von 47 % (17

von 36). Die Sensitivitäten und Spezifitäten zwischen der klinischen Befundung und dem U-Net unterschieden sich zwischen den Arbeitspunkten nicht signifikant ($P > 0,99$).

Der Dice-Wert zwischen der manuellen und der U-Net erstellten Prostatasegmentierung lag bei 0,89, der Dice-Wert zwischen den Läsionssegmentierungen lag bei 0,35.

Die gemeinschaftliche Detektion einer Läsion zwischen Prostate-Imaging-Reporting- and-Data-System-Wert ≥ 4 und U-Net am Arbeitspunkt $\geq 0,33$ verbesserte den Positiv Prädiktiven Wert von 48 % (28 von 58) auf 67 % (24 von 36) ($P = 0,01$), dabei verblieb der Negativ Prädiktive Wert unverändert bei 83 % (25 von 30 zu 43 von 52) ($P > 0,99$).

Deep-Learning-basierte Ansätze haben in der jüngeren Vergangenheit bereits eindrucksvoll ihr zukunftsweisendes Potential demonstriert, bei spezifischen Aufgaben eine dem Menschen ebenbürtige oder sogar eine bessere Leistung erbringen zu können. Die Qualität der verwendeten Kohorte ist von einem hohen Standard und nahe an der klinischen Realität. Die in dieser Studie verwendeten Materialien und Methoden sind damit als geeignet zu bewerten, die initiale Fragestellung zu beantworten. Die Ergebnisse sind vergleichbar und kongruent mit zuvor durchgeführten und auf diese Studie folgende Arbeiten, wobei sie eine hohe klinische Relevanz besitzen.

In der Zusammenschau sind die Ergebnisse ein wichtiger Schritt zu der Erkenntnis, dass ein Deep-Learning basierter Ansatz eine Prostatakarzinomdiagnose auf MRT- Bildern stellen kann, wobei in der retrospektiven Forschungsfragestellung eine der menschlichen MRT- Interpretation vergleichbare Leistung festgestellt werden konnte. .