

Lars Dominik Piskorski

Dr. med.

Künstliche Intelligenz zur Prädiktion der Malignitätswahrscheinlichkeit pulmonaler Rundherde: Vergleich und Validierung des Lung Cancer Predictor Convolutional Neural Network Scores in unterschiedlichen Patientenkohorten

Fach/Einrichtung: Radiologie

Doktorvater: Prof. Dr. med. Claus Peter Heußel

Das Lungenkarzinom ist eine der führenden Krebserkrankungen und die häufigste krebisbedingten Todesursachen in Deutschland. Die späte Diagnosestellung des Lungenkarzinoms ist eine Ursache für die geringe 5-Jahresüberlebensrate von 19% bis 23%. Eine Möglichkeit die Prognose des Lungenkarzinoms zu verbessern, stellen Lungenkrebs-Screening-Programme dar, die in Deutschland allerdings noch nicht etabliert sind. Eine weitere Möglichkeit stellt die korrekte Beurteilung von zufällig entdeckten, sogenannten inzidentellen pulmonalen Rundherden dar. Zur Beurteilung der Rundherdmalignität werden häufig klassische statistische Modelle wie der Lung-RADS® Score oder der Brock-Modell Score verwendet. Mit dem LCP-CNN (*Lung Cancer Predictor Convolutional Neural Network*) Score der Firma Optellum® steht das erste künstliche Intelligenz basierte Programm zur Malignitätsbeurteilung pulmonaler Rundherde mit einer FDA 510(k) Zulassung zur Verfügung. Ziel dieser Studie war es, den LCP-CNN Score an einem klinisch orientierten Datensatz und in unterschiedlichen Subkohorten, hinsichtlich Patientenzusammensetzung und pulmonaler Grunderkrankung, zu validieren und mit dem Brock-Modell Score und dem Lung-RADS® Score zu vergleichen.

In dieser retrospektiven Validierungsstudie wurden pulmonale Rundherde mit dem LCP-CNN Score, dem Brock-Modell Score und dem Lung-RADS® Score analysiert. Der Datensatz umfasste Thorax-CTs und klinische Daten von 297 Patienten mit 519 pulmonalen Rundherden mit einer Größe zwischen 5mm und 30mm. Von 519 Rundherden erfüllten 422 Rundherde die zuvor definierten Ein- und Ausschlusskriterien. 105 Rundherde waren maligne und 317 Rundherde waren benigne. Auf Grundlage der Gesamtkohorte wurden drei Sub-Kohorten gebildet. Eine Screening-Subkohorte umfasste Patienten, die die Einschlusskriterien für ein Lungenkrebs-Screening erfüllen würden (n=148). Eine Emphysem-Subkohorte umfasste Patienten mit einem diagnostizierten Lungenemphysem (n=136). Eine ILD-Subkohorte umfasste Patienten mit einer diagnostizierten interstitiellen Lungenerkrankung oder einer interstitiellen Lungenabnormalität (n=47).

In der Gesamtkohorte zeigte sich der LCP-CNN Score mit einer AUC (area under the curve) von 91,5% dem Brock-Modell Score mit einer AUC von 88,0% (p=0,022) und dem Lung-RADS® Score mit einer AUC von 82,9% (p<0,001) signifikant überlegen. In der Gesamtkohorte zeigte der LCP-CNN Score eine Sensitivität von 97,1% und eine Spezifität von 55,2%. Der Brock-Modell Score zeigte eine Sensitivität von 91,4% und eine Spezifität von 64,4%. Der Lung-RADS® Score zeigte eine Sensitivität von 91,4% und eine Spezifität von 63,7%. Auch in der Screening-Subkohorte zeigte sich der LCP-CNN Score mit einer AUC von 92,7% dem Brock-Modell Score mit einer AUC von 87,8% (p=0,016) und dem Lung-RADS® Score mit einer AUC von 82,1% (p<0,001) signifikant überlegen. In der Emphysem-Subkohorte zeigte der LCP-

CNN Score mit 90,8% eine leicht niedrigere AUC als in der Kontroll-Subkohorte mit 92,4%, allerdings ist dieser Unterschied mit $p=0,587$ nicht signifikant. In der ILD-Subkohorte zeigte der LCP-CNN Score mit 92,0% eine leicht höhere AUC als in der Kontroll-Subkohorte mit 91,5%, ebenfalls ohne statistische Signifikanz mit $p=0,876$.

Diese Arbeit hat die Erkenntnisse vorheriger Arbeiten bestätigt, die zeigten, dass der LCP-CNN Score dem Brock-Modell Score überlegen ist. Darüber hinaus konnte erstmalig gezeigt werden, dass der LCP-CNN Score auch dem Lung-RADS® Score überlegen war. Zusätzlich haben die Ergebnisse dieser Arbeit gezeigt, dass das Vorhandensein eines Lungenemphysems oder einer interstitiellen Lungenerkrankung bzw. einer interstitiellen Lungenabnormalität die Malignitätsprädiktion des LCP-CNN Scores nicht signifikant beeinflussen. Mögliche Limitationen dieser Arbeit umfassten unter anderem die monozentrische Rekrutierung von Patienten an der Thoraxklinik des Universitätsklinikums Heidelberg. Außerdem zeigte die ILD-Subkohorte eine geringe Rundherdzahl ($n=80$). Der Brock-Modell Score und der Lung-RADS® Score wurden zudem an inzidentellen Rundherden eingesetzt, obwohl diese primär für den Einsatz in Lungenkrebs-Screening-Studien vorgesehen sind.