

1. TEIL: KURZBERICHT

1.1 Ursprüngliche Aufgabenstellung

Die Thorax-Bildgebung mit Hilfe der CT ist eine der häufigsten radiologischen Untersuchungen in Deutschland und weltweit. Insgesamt wurden 2009 ca. 800.000 Untersuchungen jährlich in Deutschland durchgeführt (Arztreport der BARMER 2011; jüngere Erhebungen sind nicht bekannt). Gemäß dem Bundesamt für Strahlenschutz (BfS) nimmt die Anzahl der CT-Untersuchungen in Deutschland schnell zu. Im Zeitraum von 2007 bis 2016 betrug der Zuwachs 45%, das sind Raten von im Mittel 5% pro Jahr. Damit schätzen wir die Gesamtzahl der Thorax-CT-Aufnahmen, die 2020 in Deutschland angefertigt werden, auf ca. 1,3 Millionen ab. Die Thorax-CT ist vor allem bei der Diagnose und Therapieüberwachung von Bronchialkarzinomen und Tumor-Metastasen etabliert. Zudem hat die Coronavirus-Pandemie gezeigt, dass die Thorax-CT-Bildgebung auch bei COVID-19-Patienten eine zentrale Bedeutung für die Diagnose und Therapieüberwachung hat. Mit CT-Verlaufskontrollen können Progression von Lungenerkrankungen, die das Lungenparenchym verletzen oder verändern, mit Hilfe von objektivierbaren Parametern überwacht und auch bereits leichte Veränderungen früh erkannt werden. Letzten Endes bieten Verlaufskontrollen die Möglichkeit, auch kleinste Therapieeffekte und Nebenwirkungen zu objektivieren und damit eine frühe Optimierung der Therapie vorzunehmen. Im klinischen Alltag wird nur ein kleiner Teil des Potentials ausgeschöpft, da keine routinetauglichen Tools und Services zur genauen, umfassenden Erfassung und Dokumentation aller krankheitsrelevanten Parameter und quantitativen Marker existieren. Die Aufgabenstellung des F&E-Vorhabens, das von den Kooperationspartnern jung diagnostics GmbH (JDX), Fraunhofer MEVIS (FME) und Universitätsmedizin Mainz (UMM) durchgeführt wurde, bestand darin, die Entwicklung einer vollautomatisch arbeitenden Bildverarbeitungstechnologie zur Erkennung kleinster anatomischer und pathologischer Veränderungen des Lungenparenchyms auf Thorax-CT-Aufnahmen voranzutreiben, die den Anforderungen in der Patientenversorgung vollumfänglich entspricht. Es sollten vorhandene Verfahren zur Detektion und Quantifizierung von Pathologien zusammengeführt, automatisiert und weiterentwickelt werden. Die Aufgabenstellung des Teilvorhabens von JDX bestand in der automatisierten Erkennung von unspezifischen Abweichungen von der Norm (Anomalien) und der Entwicklung von Verfahren zur Quantifizierung von Pathologien des Lungenparenchyms. Weiterhin sollten Konzepte zur Befundunterstützung entwickelt werden, die aufbauend auf vollautomatisiert durchgeführten Bildanalysen, die Dauer der Befundung signifikant reduzieren helfen.

1.2 Ablauf des Vorhabens

Ein für die Anwendung von Deep Learning (DL) Verfahren ausreichend großer Trainingsdatensatz für die Lungenrundherddetektion und die Quantifizierung der Lungenrundherd-Verände-

rung wurde erarbeitet (Arbeitspaket #1 (AP1); Teilziel #1). Eine auf Deep Learning (DL)-Technologien basierende robuste Detektion von Anomalien (mit Fokus auf Lungenrundherde) wurde entwickelt (AP2; Teilziel #2). Im Rahmen von AP4 wurden verschiedene technologische Ansätze für eine longitudinale Verlaufsanalyse von Rundherden auf Basis eines neuartigen DL-basierten Registrierungsverfahrens wurden grundlegend untersucht und auf das Problem angepasst (Teilziel #3). Basierend auf den technologischen Ergebnissen auf AP2 und AP4 wurden im Rahmen von AP5.1 ein Visualisierungskonzept entwickelt, um den befundenen Radiologen die automatisch detektierten und quantifizierten Veränderungen der Lungenrundherde optimal darzustellen (AP5.1). Zusätzlich wurde ein Algorithmus entwickelt, der auf Basis der gefundenen Lungenrundherde und der quantifizierten Lungenrundherd-Veränderung einen automatisch generierten Befundtext als radiologischen Vorbefund erzeugt (AP5.2). Ein Demonstrator für die Lungenherdetektion und die Quantifizierung der Lungenherdveränderung, der auch die technologischen Entwicklungen der Partner aufgenommen hat, wurde erfolgreich entwickelt und in die kommerzielle Bildverarbeitungspipeline von JDX (Biometrica) als Add-on Module integriert (AP5.3). Im Rahmen von AP6 wurde ein Pilotversuch mit dem Demonstrator unter Real-World Bedingungen bei zwei Pilotkunden vorbereitet und durchgeführt. Erst dadurch konnte der Impact der voll-automatisch generierte Befundvorschläge auf den Workflow in Pilotzentren untersucht werden. Ein Proof-of-Concept konnte damit erfolgreich etabliert werden (Teilziel #4).

1.3 Wesentliche Ergebnisse

Im Rahmen dieses Förderprojekts wurden grundlegende Bildverarbeitungstechnologien zur automatischen Analyse und Befundung von CT-Untersuchungen der Lunge erforscht und bis zur Anwendungsreife entwickelt. Das wesentliche Ergebnis des Teilvorhabens von JDX ist eine vollautomatisch arbeitende Bildverarbeitungs-Pipeline (SPIRABENE-System) auf Basis von DL-Technologien zur Erkennung kleinster pathologischer Veränderungen auf CT-Aufnahmen der Lunge. Das SPIRABENE-System besteht aus zwei Modulen: (1) einer Detektion von kleinsten (unspezifischen) pathologischen Veränderungen im Lungenparenchym bei Vorhandensein einer einzigen Aufnahme des Patienten (Querschnitt-Analyse), die eine automatische Differenzierung von normalen (altersbedingten) und abnormalen Veränderungen durchführt, und (2) einer Komponente für die Verlaufsanalyse, die bei Vorhandensein einer Voruntersuchung vom selben Patienten die Veränderung zwischen zwei Untersuchungen im Bereich des Lungenparenchyms quantifiziert. Als erster Anwendungsfall wurde das SPIRABENE-System für die Detektion von Lungenrundherden auf Basis von Niedrig-Dosis CT-Untersuchungen optimiert und validiert. Sie soll nach erfolgter Medizinproduktzulassung in der Lungenkrebs-Früherkennung eingesetzt werden. Das SPIRABENE-

System erfüllt vollumfänglich die Anforderungen an eine Software zur computerassistierten Detektion, wie sie für das nationalen Lungenkrebs-Früherkennungs-Programm¹ definiert wurden. Da sie im Vergleich zu konkurrierenden Ansätzen nicht nur auf Rundherde, sondern auch bzgl. normalen Strukturen, wie Lungenlappen, Bronchien und Arterien, trainiert wurde, ist sie in der Routineversorgung den Konkurrenzprodukten überlegen, in der die Bildqualität oft suboptimal und Störgrößen, wie Bildartefakte, häufig vorhanden sind.

¹ Verordnung über die Früherkennung von Lungenkrebs bei rauchenden Personen (LuKrFrühErKV) vom 17. Mai 2024: <https://www.recht.bund.de/bgbl/1/2024/162/VO.html>

2. TEIL: EINGEHENDE DARSTELLUNG

2.1 Durchgeführte Arbeiten

Die detaillierte Darstellung der im Rahmen des Teilprojekts mit dem Titel „Quantifizierung und Konzepte zur Befundunterstützung“ von JDX geleisteten Arbeiten erfolgt in der Reihenfolge der Arbeitspakete, so wie sie im Arbeitsplan der Teilvorhabenbeschreibung ursprünglich definiert worden sind.

Arbeitspaket	Task	Monate	Aufwand in PM			Monate ab Projektbeginn											
			JD	FM	UK	Jahr 1				Jahr 2				Jahr 3			
			X Σ5 7	E Σ2	T Σ8	Q1 1-3	Q2 4-6	Q3 7-9	Q4 10-12	Q1 13-15	Q2 16-18	Q3 19-21	Q4 22-24	Q1 25-27	Q2 28-30	Q3 31-33	Q4 34-36
AP1 (Lead: UKT) Daten und Annotationen (Σ3PM/JDX 2 PM/UKT 1 PM)	1.3 Normal-Daten	1-12	2	-	1												
	Meilensteine			M2													
AP2 (Lead: FME) Pathologie- detektion	2.4 Anomalie-Detektion I	1-12	7	-	-												
	2.5 Anomalie-Detektion II	10-21	6	-	-												
	Meilensteine			M X													
AP4 (Lead: JDX) Quantifizierung (Σ17PM/JDX 17 PM)	4.1 Verm. Dichteunterschiede	15-24	6	-	-												
	4.2 Externe Einflussfaktoren	18-30	6	-	-												
	Meilensteine			MX M3 M4													
AP5 (Lead: JDX) Strukturierte Befundung und klinischer Prototyp (Σ22PM/JDX 13 PM/FME 2 PM/UKT 7 PM)	5.1 Visualisierungskonzept	22-30	5	-	-												
	5.2 Radiologische Befunde	12-24	6	-	-												
	5.3 SW-Demonstrator	21-33	6	-	-												
	Meilensteine			M X M3 M4													
AP6 (Lead: UKT) Klinische Evaluierung (Σ0PM)	6.1 Retrospektive Evaluierung	27-33	4	-	2												
	6.2 klin. Workflow	30-33	4	-	1												
	6.3 Pilotstudie	32-36	5	2	4												
	Meilensteine			M4													

AP1.3 Erstellen einer CT-Schichtbild-Sammlung nicht-pathologischer Regionen (Normal-Datenbank)

Mittels der öffentlichen LIDC Daten (<https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254>) konnten wir die Komplexität der Anomalie-Detektion auf Thorax CT Bildern erfassen. Auf jeder Schicht des Datensatzes haben vier Experten die Anomalien markiert. Die Schichten eines CT-Datensatzes, die ohne Markierung sind, wurden als normal klassifiziert. Auf diese Weise wurden ca 200.000 normale Schichten identifiziert und in einer Datenbank organisiert.

AP2.4 Anomalie-Detektion I: Entwicklung eines Autoencoders für Normal-Darstellung von CT-Daten ohne anormale Veränderungen.

Bislang wurden Deep-Learning-basierte Verfahren zur Rekonstruktion von normalen Anatomien erfolgreich in relativ starren und wenig deformierbaren Strukturen, zum Beispiel im Bereich des Schädels und Hirnparenchyms angewendet. Wir haben uns in AP2.4 ausführlich mit der Möglichkeit beschäftigt, Autoencoder für Normal-Darstellungen von CT-Daten ohne abnormale Veränderungen auch bei elastischen Strukturen einzusetzen. Zum anderen haben wir herausgearbeitet, dass ein vollständig unüberwachter (unsupervised) Ansatz, der jegliche abnormale Veränderungen detektiert, vermutlich zu unspezifisch ist, da es eine zu große Anzahl verschiedener abnormaler Veränderungen gibt. Beispiele sind Lungenrundherde oder Lungennodule, Lungentumore, Fibrose, Narbengewebe, Wassereinlagerungen, kollabierte Teile der Lunge etc. Viele dieser Veränderungen sind unkritisch. Häufig sind nur besondere Arten von Veränderungen von Interesse. Deshalb haben wir uns dazu entschlossen, die Normal-Darstellung der CT-Daten ohne abnormale Veränderungen, statt als eigenständigen Ansatz, als Unterstützung zu einem klassischen supervised Ansatz zu verstehen, der sich nur auf eine spezifische Anomalität fokussiert. Aufgrund dieser Entscheidung gehören die Arbeiten an diesem supervised Ansatz nicht mehr zum Teilarbeitspaket AP2.4, sondern gehören inhaltlich zu AP2.5 und werden dort beschrieben.



Abbildung 1. Beispiel der automatischen Segmentierung der Lungenlappen. Die fünf Farben spiegeln die fünf Lungenlappen wider.

Wesentlich für AP2.4 ist aber nach unserer Erfahrung, dass automatische Verfahren zur Detektion von abnormalen Veränderungen sehr davon profitieren, wenn zusätzlich gesunde anatomische Strukturen mit in das Training einbezogen werden. Konsequenterweise haben wir uns daher in

diesem Arbeitspaket mit der Erkennung (Segmentierung) von gesunden Anatomien, die vermutlich einen signifikanten Einfluss auf die Detektion von Anomalien haben werden, auf Basis öffentlich-verfügbarer Daten beschäftigt. Dabei konnten wir Modelle zur Segmentierung der fünf Lungenlappen, der Arterien (Herz und Lunge) und der Atemwege (Bronchen und Luftröhre) erstellen.

Die Segmentierung der Lungenlappen erfolgte mittels des Totalsegmentator-Datensatzes (<https://github.com/wasserth/TotalSegmentator>). Dies ist der einzige öffentliche Datensatz, der schon Segmentierungen der Lungenlappen enthält. Allerdings wurden die Annotationen mit einem Fokus auf Quantität statt Qualität erstellt, weswegen ein großer Teil unserer Arbeit aus der manuellen Nachbearbeitung qualitativ schlechter Annotierungen bestand. Die restlichen Daten wurden zum Training und zum sorgfältigen Fine-Tuning unseres Modells genutzt. In diesem Zuge wurden eine Vielzahl von Modellen trainiert und sowohl quantitativ mittels bekannter Metriken (z.B. Dice Similarity Score (DSC), Hausdorff-Distanz) als auch visuell qualitativ überprüft. Das finale Modell war in der Lage, robust Segmentierungen auf einer Vielzahl an heterogenen Scans aus Versorgung zu erstellen. Das Modell segmentiert die fünf Lungenlappen mit einer Median-Genauigkeit von 97.8% im Mittel über die fünf Strukturen und trennt Lungengewebe von restlichem Gewebe mit einer Median-Genauigkeit von 99.0% (binäre Segmentierung). Die Fehler sind Ungenauigkeiten an den Konturen der Lungen oder beim Übergang zwischen zwei verschiedenen Lappen, Inkonsistenzen beim Ein- und Ausschluss von großen Pathologien am Rand der Lunge und falsch positive Annotationen außerhalb der Lunge. All diese Fehler sind auf den verwendeten Datensatz zurückzuführen und werden durch Aufbessern in Zukunft beseitigt. Ein typisches Beispiel für die automatische Lungenlappensegmentierung ist in Abbildung 1 zu finden. Die automatisch generierten Segmentierungen der Lungenlappen wurden im Verlauf der erstellten Pipeline verwendet, um eine Lokalisation der detektierten Lungennodule anzugeben, also um anzugeben, in welchen Lungenlappen sich die Nodule befinden. Obwohl die Segmentierung der Lungenlappen eine natürliche Ungenauigkeit beim Übergang zwischen zwei Lappen zeigte, hat sich die Qualität dieser Lokalisation als hinreichend erwiesen. Aus diesem Grund haben wir uns dazu entschieden, die bestehende Lungenlappensegmentierung als erfolgreich zu betrachten und keine weiteren Versuche zu unternehmen, letzte Fehler zu beheben. Dies würde die manuelle Segmentierung von Lungenlappen benötigen, die aufgrund der Größe der Strukturen zeitliche zu aufwändig wäre und somit der Aufwand den Nutzen nicht gerechtfertigen würde.

Die Segmentierung der Arterien und Atemwege erfolgte mittels CT Daten aus öffentlichen „segmentation challenges“. In diesen werden CT Bilder und Segmentierungen veröffentlicht und Teams dazu aufgefordert, innerhalb weniger Monate einen möglichst präzisen Algorithmus zu trainieren. Unsere Lösung für automatische Erkennung von Arterien und Atemwegen wurde durch das Zusammenführen von zwei separaten Challenge-Datensätze (<https://atm22.grand->

challenge.org/) für die jeweiligen Aufgaben erfüllt. Die ATM22 (<https://atm22.grand-challenge.org/>) und Parse2022 (<https://parse2022.grand-challenge.org/>) challenges beinhalten Datensätze zur Segmentierung der Bronchen bzw. Arterien. Auf den benannten Daten wurden sorgfältig eine Vielzahl an Modellen trainiert. Das Training hat ergeben, dass es von Vorteil ist, ein Modell zu verwenden, das beide Strukturen auf einmal segmentiert anstatt zwei separate Modelle zu verwenden. Die Herausforderung in diesem Training bestand darin, dass bei beiden Datensätzen je nur eine der beiden Strukturen als manuelle Grundwahrheit vorliegt. Deshalb mussten die bestehenden Trainingspipelines angepasst werden, um dies berücksichtigen zu können. Das finale Modell erzielte bei den Atemwegen einen DSC Wert von 80.1% und bei den Arterien von 78.6%. Das Modell erstellte robuste Segmentierungen auf heterogenen Scans aus der Versorgung die visuell als hinreichend gut bewertet wurden. Abbildung 2 zeigt ein solches Beispiel.

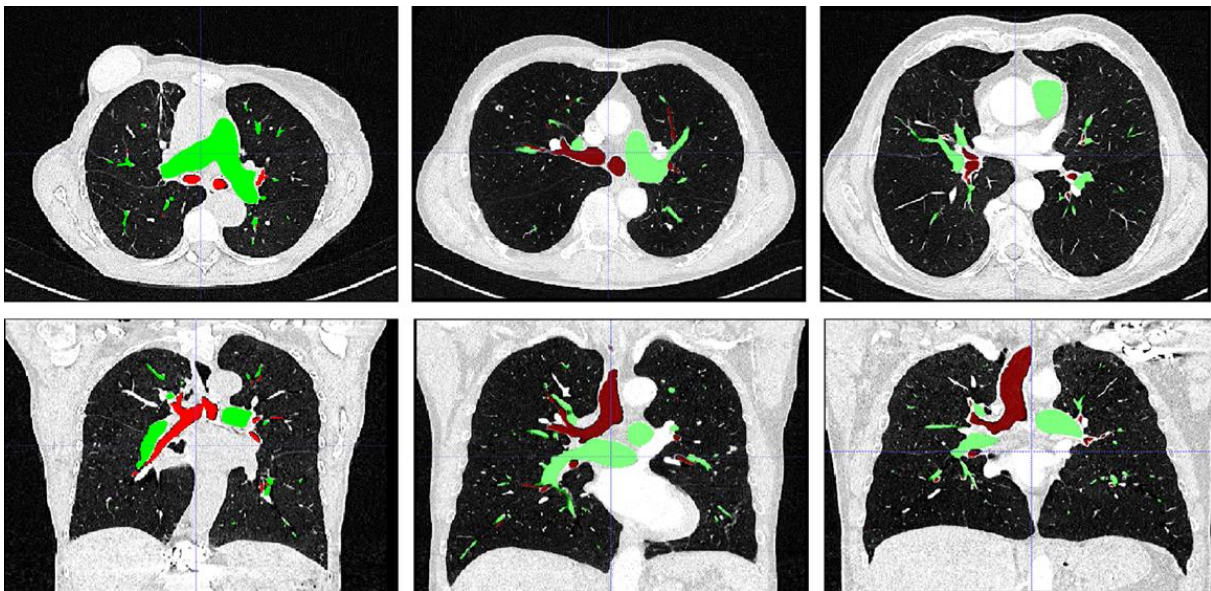


Abbildung 2. Beispiel einer vollautomatischen Segmentierung von Arterien (grün) und Atemwegen (rot), die in diesem Projekt entwickelt wurde. Das Arterien-System umfasst insbesondere die Herzarterie (mittig), aber auch kleine Gefäße bis in die Lungen. Die Atemwege bestehen innerhalb der Lunge aus den Bronchien und der Luftröhre. Durch die gleichzeitige Erkennung dieser Anatomien und der Anomalien konnte die Rate an falsch positiven Annotationen drastisch verringert werden, wie in AP2.5 erläutert.

AP2.5 Anomalie-Detektion II: Detektion von Anomalien mit Hilfe des Autoencoders und Weiterentwicklungen

In diesem Arbeitspaket haben wir uns die am häufigsten vorkommende Anomalie im Lungenparenchym, nämlich den Lungenrundherden (Lungennodulen), fokussiert. Diese spielen eine zentrale Rolle bei der Früherkennung von Lungenkrebs. Die Entscheidung für diese Zielstruktur wurde getroffen, da aktuell Vorbereitungen auf ein nationales Lungenkrebscreening getroffen werden. Da die ersten Vorversuche ergeben haben, dass ein genereller Angang über Autoencoder

oder Diffusionsmodelle zu risikoreich erscheint, haben wir einen „Supervised“-Ansatz gewählt, um einen in der Routineversorgung anwendbaren Algorithmus zur Anomalie-Detektion zu realisieren. Die automatische Erkennung von Lungenrundherden ist eine große Herausforderung, da diese oft sehr klein sind und Ähnlichkeiten zu Abzweigungen von Lungengefäßen oder rundem Narbengewebe aufweisen können. Um den Anforderungen an die Praxis muss eine hohe Sensitivität erreicht werden.

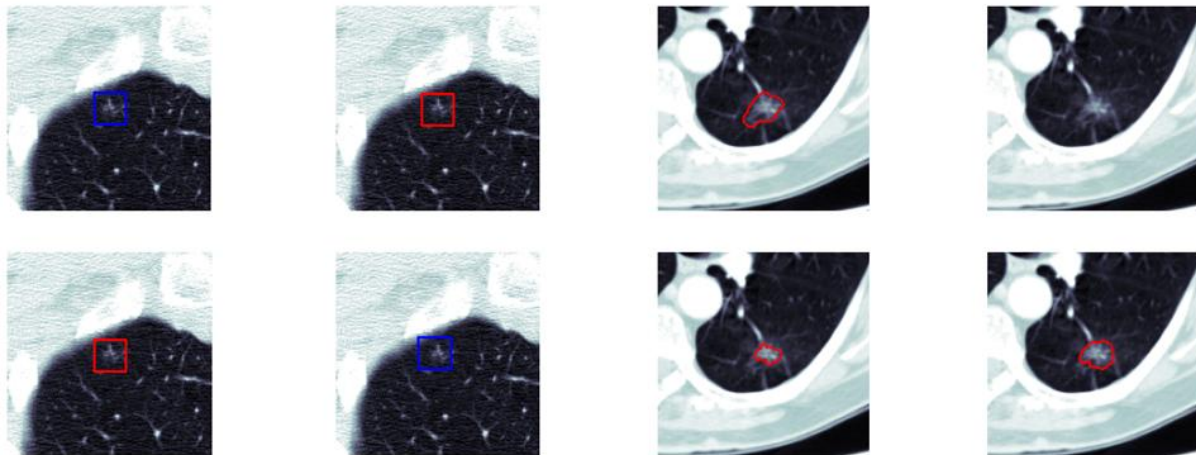


Abbildung 3. Beispiele für die Inter-reader Variabilität bei Radiologen. Die linken beiden Spalten zeigen Anatomien die kleiner als 3mm sind und von allen vier RadiologInnen markiert wurden, allerdings in zwei Fällen als Lungenknötchen (rot) und in zwei Fällen als nicht-Lungenknötchen (blau). Die rechten beiden Spalten zeigen manuelle Segmentierungen eines Lungenknötchens mit Größe von mehr als 3mm. Hier wird deutlich, dass die Konturierung hoch variabel.

Auf den in AP1.3 aufbereiteten öffentlichen Daten sind manuelle Annotationen von insgesamt vier Radiologen vorhanden, die wir zur Messung der Abweichung unter jenen (inter-reader Variabilität) nutzen konnten. Beispiele für die Inter-reader Variabilität zwischen Radiologen auf dem verwendeten Datensatz sind in Abbildung 3 zu sehen. Beispielsweise wurde eine mittlere Sensitivität von 66.5% für Lungenknötchen mit einem Durchmesser von weniger als 3 mm erreicht, was auf eine hohe Komplexität zur Erkennung von Anomalien dieser Größe schließen lässt. Wir haben zunächst Modelle zur Segmentierung von Lungenknötchen trainiert, die auf einem Standard 3D U-Net basieren. Dafür haben wir das weit verbreitete nnU-Net Framework als baseline-Modell verwendet und nach und nach einzelne Parameter des Modelles angepasst, um eine Verbesserung der Performance zu erreichen. Wie unser Modell für die Segmentierung der Lungenlappen, betrachtet dieses Netzwerk nicht das gesamte 3D CT-Bild, sondern nur einen kleinen Ausschnitt von diesem. Um ein vollständiges 3D CT-Bild auszuwerten, wird das vollständige Bild in mehrere sich überlappende Ausschnitte (Batches) unterteilt, einzeln evaluiert und schließlich wieder zusammengefügt. Das Netzwerk wurde so trainiert, dass nicht nur eine Segmentierung der Lungen- nodule, sondern auch die Segmentierung der Lunge (Lungenlappen binarisiert), von Arterien und

Atemwegen erfolgt. Die Fehleranalyse unserer trainierten Modelle zur Erkennung von Lungennodulen zeigte, dass besonders Abzweigungen von Gefäßen und kleine Wölbungen am Rand der Lunge falsch positive markiert wurden. Insgesamt zeigte sich, dass dieser Ansatz zu einem ausgegogenen Verhältnis von falsch positiven und falsch negativen Annotationen führte. Dies ist ein wünschenswertes Ergebnis, wenn die Methode alleinstehend verwendet werden soll, um Scans zum Beispiel im Rahmen einer Lungenkrebs-Früherkennung zu befunden. Allerdings ist das in Deutschland beschlossene nationale Lungenkrebscreening-Programm so geplant, dass der Algorithmus als Unterstützung für den Radiologen dienen soll, der final die Befundung vornimmt. Das bedeutet in der Praxis, dass der Algorithmus möglichst den Befunder auf mögliche Nodule hinweisen soll, die vorher übersehen wurden. Somit ist es wichtig, eine möglichst hohe Sensitivität zu erreichen, falsch positive Annotationen sind in diesem Szenario nicht so schwerwiegend wie falsch negative. Diese Anforderungen sind mit dem ersten Ansatz nur schwer umsetzbar, da dieser keine Flexibilität bietet um dynamisch zwischen wenigen falsch positiven und wenigen falsch negativen Funden wechseln zu können. Daher haben wir weitere Ansätze untersucht. Der finale und erfolgreiche Ansatz besteht aus einer Pipeline, die drei Schritte umfasst (i) einer sehr sensitiven Segmentierung von potentiellen Lungennodulen, (ii) einer binären Klassifizierung der Funde in echt und falsch-positive und (iii) einer anschließenden Korrektur der Kontouren des Noduls. Für den ersten Schritt haben wir erneut die Segmentierungen von Lungenlappen, Arterien und Atemwegen als zusätzliche Klassen verwendet. Um eine möglichst hohe Sensitivität zu erreichen wurden alle Nodule für das Training verwendet, die von mindestens einem Experten annotiert worden sind. Zusätzlich haben wir alle Strukturen die als „nicht-Nodule“ markiert worden sind ebenfalls zum Training verwendet. Da diese nur als Landmarke markiert, aber nicht segmentiert wurden, haben wir ein einfaches Netz trainiert, das als Eingabe einen kleinen Bildausschnitt erhält und eine Nodul-ähnliche Struktur in der Mitte segmentiert. Das resultierende Modell weist eine hohe Sensitivität auf, produziert allerdings auch eine große Menge an falsch positiven Funden. Wir konnten beobachten, dass das Netzwerk ca. 96% der von Experten annotierten Noduli findet, dass allerdings mehr als 90% der Funde falsch positiv sind. Das zweite Netzwerk der Pipeline ist ein Klassifizierung-Netzwerk. Als Input fungiert ein kleiner Bildausschnitt, in dessen Zentrum sich ein potentielles Nodul befindet. Zusätzlich wird dem Netzwerk die binäre Maske des potentiellen Noduls übergeben. Das Netzwerk wird so trainiert, dass es die relative Anzahl an Experten schätzen soll, die den Fund als Nodul einschätzen, also 0 bzw. 1 falls kein bzw. alle Experten den Fund als Nodul betrachten würden. Im Gegensatz zu anderen Ansätzen bei dem nur ein binärer Wert (0 oder 1) ausgegeben wird, liegen bei unserem Ansatz kontinuierliche Scores vor, die angeben, mit welcher Sicherheit ein Fund als Nodul betrachtet werden kann. Dies hat den Vorteil, dass der Schwellwert, nach welchem entschieden wird, ob ein Nodul im Bericht erscheinen soll oder nicht, dynamisch gewählt werden kann, um ein gewünschtes Verhältnis von falsch positiven zu falsch negativen Funden zu erreichen. Im letzten Schritt werden die Kontouren der bestätigten Nodule

korrigiert. Dies ist notwendig, da im ersten Schritt, der sensitiven Segmentierung der Nodule, nicht nur zu viele Annotationen vom Netzwerk getätigt werden, sondern diese auch noch überzeichnet werden. Dies bedeutet eine systematische Überschätzung des Volumens der Nodule, die korrigiert werden muss. Wie im vorherigen Schritt erhält das Netzwerk einen Ausschnitt des CT Bildes und die binäre Maske des Noduls als Eingabe, dieses Mal mit dem Ziel, die Kontouren des Noduls möglichst genau zu bestimmen.

AP4.1 Vermessung von lokalen Dichteunterschieden des Lungengewebes (im Querschnitt und longitudinal); Störeinflüsse von Atmungszuständen.

Die Aufgaben in diesem Paket wurden in zwei Unteraufgaben unterteilt. Zum einen betrachten wir in diesem Arbeitspaket die Vermessung der Veränderung im Volumen einzelner Lungenrundherde zwischen zwei Aufnahmen. Zum anderen haben wir uns damit beschäftigt, unsere Lösungen robust gegen Störeinflüsse, wie zum Beispiel Atmungszustände, zu trainieren. Letzteres haben wir durch das Verwenden von Methoden der Datenaugmentierung erreicht. Datenaugmentierung ist eine Technik, bei der der Input des Modells während des Trainings absichtlich verzerrt wird, um mehr Trainingsbeispiele zu erzeugen und das Modell invariant gegen gewissen Störungen zu trainieren. Beispielsweise lernt dabei das Modell, dass sich bei einer 20 Grad Rotation um den Ursprung die Zielsegmentierung ebenfalls um 20 Grad dreht, während sich diese bei einer Verrauschung des Bildes nicht ändert. Solche Techniken sind essentiell, um das Netzwerk für die Verwendung auf Daten, die den Trainingsdaten nicht ähneln, vorzubereiten. In diesem Arbeitspaket haben wir insbesondere Deformationen verwendet, um das Modell sehr gezielt auf Störeinflüsse, wie Atembewegungen vorzubereiten. Insbesondere Spiegelungen, Rotationen, Verkleinerung und Vergrößerung der Bilder haben sich in dieser Anwendung als effektiv erwiesen. Die Tests ergaben, dass mit dieser Datenaugmentierungs-Methode eine ausreichende Stabilität im Vergleich von Bildern mit leicht unterschiedlichen Atemzuständen erreicht werden konnte.

Die zeitliche Vermessung der Lungenrundherde wird im Lungenkrebscreening durch die Berechnung der Volumenverdopplungszeit (VDT) abgebildet. Diese Größe bestimmt wie sehr sich das Volumen jedes einzelnen Noduls in einer gewissen Zeitspanne verändert hat. Sie beschreibt in vielen Tagen sich das Volumen des Noduls verdoppelt hat. Zur Berechnung werden die Volumina zum ersten und zweiten Aufnahmezeitpunkt (v_0, v_1) benötigt, sowie die Anzahl an Tagen die zwischen dem ersten und zweiten Scan vergangen sind ($t_1 - t_0$). Die VDT berechnet sich dann mit folgender Formel:

$$VDT = \frac{t_1 - t_0}{\log_2(v_1) - \log_2(v_0)}$$

Im Fall eines schrumpfenden Noduls ist $VDT < 0$ und beschreibt die (negative) Anzahl an Tagen, in der sich das Volumen des Noduls halbiert. Die VDT wird dann mit in der Literatur beschriebenen Schwellwerten weiter interpretiert. Nodule mit einem positiven VDT Wert von unter 400 Tagen werden als schnell wachsend klassifiziert und haben ein hohes Krebsrisiko. Befindet sich der Werte zwischen 0 Tagen und -400 Tagen, ist das Nodul schrumpfend und hat kein Krebsrisiko. Alle anderen werden als stabil klassifiziert und haben ein geringes Krebsrisiko. Die Herausforderung bei der Bestimmung der VDT, ist die Identifikation von Nodulpaaren zwischen den zwei Zeitpunkten, d.h. zu identifizieren welches Nodul im Follow up-Bild zu einem anderen Nodul im Baseline-Bild passt. Der anatomische Bereich, die der jeweilige Scan zeigt und der Atemzustand der Lunge unterscheiden sich zwischen den Aufnahmen leicht. Deshalb wird eine elastische Registrierung benötigt, um die Positionen der Lungennodule in den zwei Aufnahmen verbinden zu können. Die von Fraunhofer-MEVIS im Rahmen dieses Verbund-Projekts entwickelte Lösung wurde für diesen Zweck getestet. Anekdotisch konnte gezeigt werden, dass die Lösung zuverlässig das Nodul im zeitlichen Verlauf identifizieren konnte. Beispiele von registrierten Bildern aus der Versorgung von berechneten Wachstumsraten sind in Abbildung 4 zu finden.

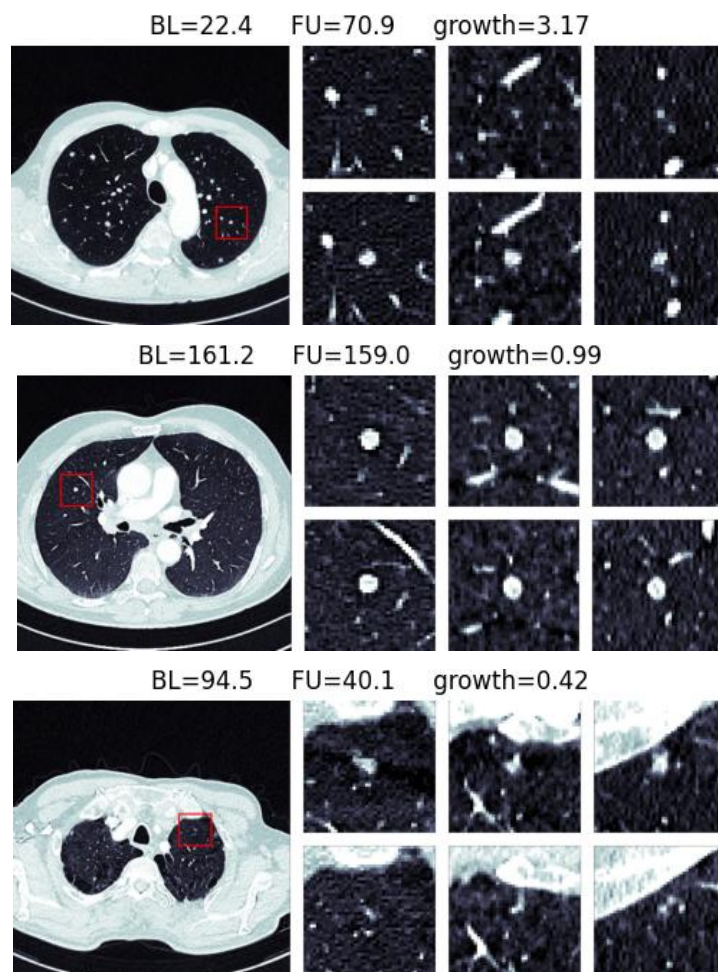


Abbildung 4. Beispiele von Lungennodulpaaren, die im zeitlichen Verlauf gewachsen (obere Zeile; $growth = 3.17$), stabil geblieben (Mitte; $growth = 0.99$) und geschrumpft sind. (untere Zeile; $growth = 0.42$). Dabei zeigen in jeder der drei Fälle die obere Zeile die drei Ansichten des Lungennodulherds

in sagittaler, koronarer und transversaler Bildorientierung der Voraufnahme und die untere Zeile den Lungenrundherd in der Verlaufsuntersuchung.

AP4.2 Behandlung von Einflussfaktoren wie Scanner- und Protokollunterschiede (Low dose-CT)

In diesem Arbeitspaket haben wir uns zunächst auf Störeinflüsse durch Bildrauschen, Bildauflösung und Unterschiede in der Bildrekonstruktion fokussiert. Diese variieren stark in der Praxis, da verschiedene Zentren unterschiedliche Scanner- und Akquisitionsprotokolle nutzen. Um die verwendeten Modelle invariant gegen solche Änderungen zu trainieren, haben wir uns auf die Erstellung verschiedener künstlicher Störmethoden fokussiert. Diese Methoden verändern die Bildqualität variabel während des Trainings und sorgen dafür, dass die resultierende Methode mit einer großen Bandbreite an Bildqualität vertraut wird. Dabei wird der Kontrast der Bilddaten verändert durch das Anwenden von Sigmoid Funktionen, die Grauwerte werden neu skaliert durch Verschiebung von Mittelwert und Standardabweichung, die Auflösung wird durch Gaussche Blurring Operationen mit variabler Stärke verringert, und zum Schluss wird additives Gauß Rauschen verwendet um das Rauschlevel zu erhöhen.

In der Praxis konnten wir beobachten, dass diese Datenaugmentierung die Robustheit der Detektion der Nodule steigert. In den Trainingsdaten befinden sich fast ausschließlich Scans die mit einer hohen Strahlendosis aufgenommen wurden und ein geringes Level an Rauschen aufweisen. Im Rahmen des Lungenkrebscreenings werden allerdings Daten mit einer geringen Strahlendosis erhoben die ein hohes Level an Bildrauschen aufweisen. Trotz dieses Unterschiedes konnten wir feststellen, dass unsere Pipeline eine hohe Performance auf den Versorgungsdaten erreichen konnte. Eine genauere Quantifizierung dessen ist in AP X.Y zu finden. Beispiele für high-resolution und low-dose CT-Scans können in Abbildung 5 gefunden werden.

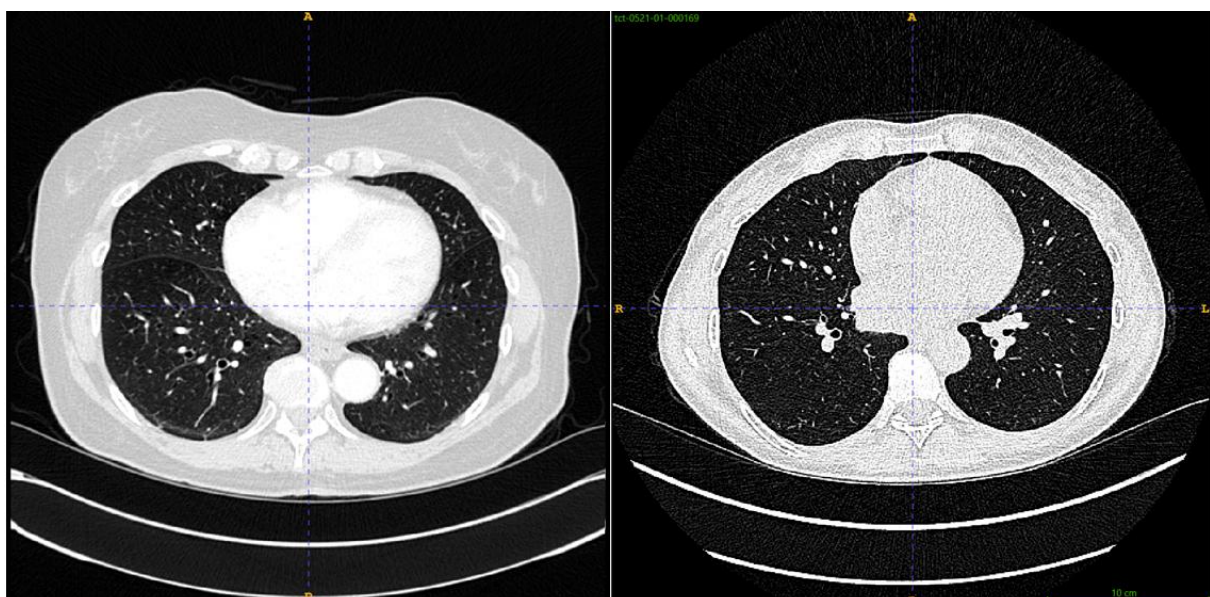


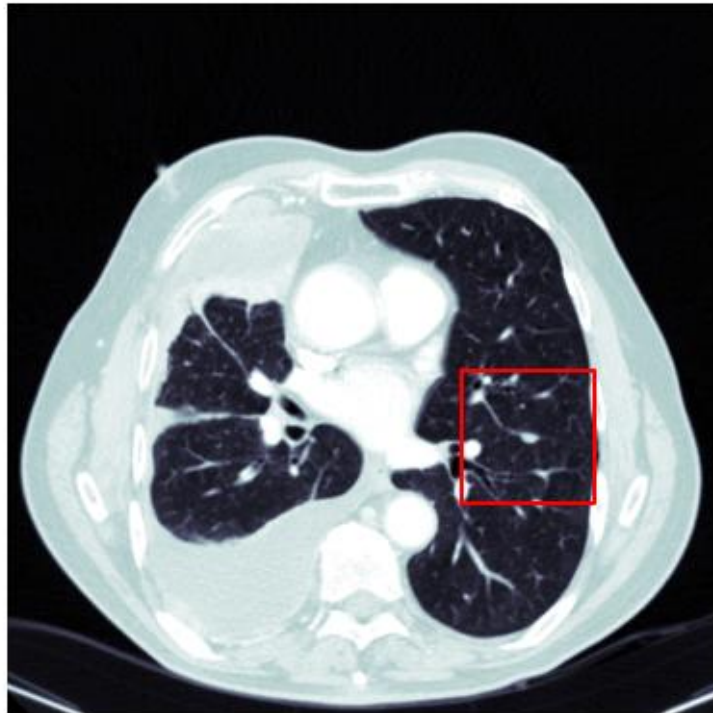
Abbildung 5. Beispiele für hochaufgelöste CT (High-resolution CT; HRCT) Trainingsbilder (links) und Niedrigdosis CT (Low-Dose CT; LDCT) Versorgungsbilder (rechts) der Lunge. Besonders in Flächen mit konstanten Grauwerten ist zu erkennen, dass der Rauschlevel bei LDCT deutlich erhöht ist. Dies erschwert die Differenzierung zwischen statistischem Rauschen und kleinen Gefäßen, bzw. Nodulen innerhalb der Lunge.

AP5.1 Entwicklung eines Visualisierungskonzepts und Darstellung der Unsicherheiten

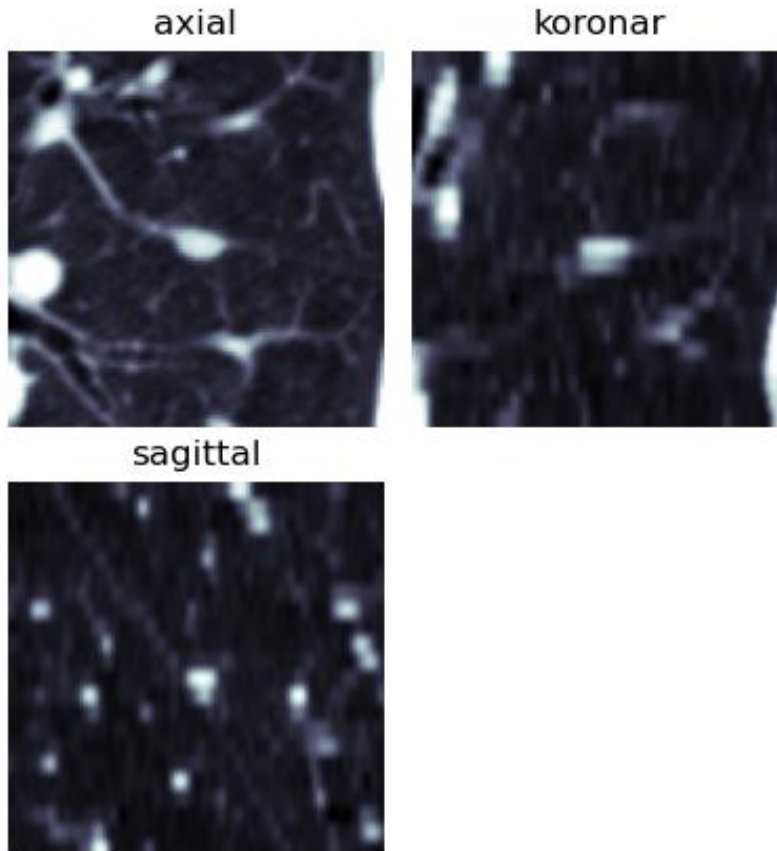
Die Visualisierung der Ergebnisse geschieht aktuell auf mehreren Ebenen (Bildorientierungen) für Fälle, in denen Lungenrundherde von der KI detektiert worden sind. Zunächst wird eine Übersicht präsentiert, in der alle Funde in der Frontalansicht als Render-Plot zusammen mit Knochen und Lungenlappen dargestellt werden:



Danach wird jeder einzelne Fund als axial-plot dargestellt. Dies ist die typische Sichtweise in der Thorax Radiologen und Radiologinnen solche Bilder befunden:



Zum Ausschluss von einem falsch positiven Fund und Einschätzung der Morphologie werden zuletzt noch andere Schnittrichtungen präsentiert:

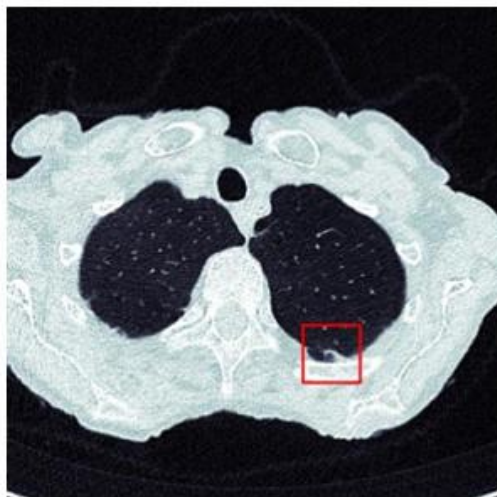


Vollständig wird die Bildinformation in einem Bericht wie folgt zusammengefasst:

Biometrica-Report

Name: Mustermann, Max
 G-Datum: 01.01.1957
 U-Datum: 14.11.2024

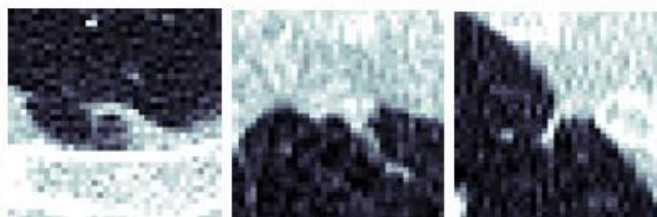
Lungenrundherd #1, linker, oberer
 Lungenlappen (Schicht = 57)



axial

koronar

sagittal



Durchmesser [mm]	8 – 10
Volumen [mm ³]	100 – 250

Für den Fall, dass Bilddaten zu einem früheren Zeitpunkt vorliegen werden diese ebenfalls dargestellt und die Angaben durch Verdopplungszeiten erweitert:

Biometrica-Report

Name: Mustermann, Max
 G-Datum: 01.01.1957
 U-Datum: 14.11.2024

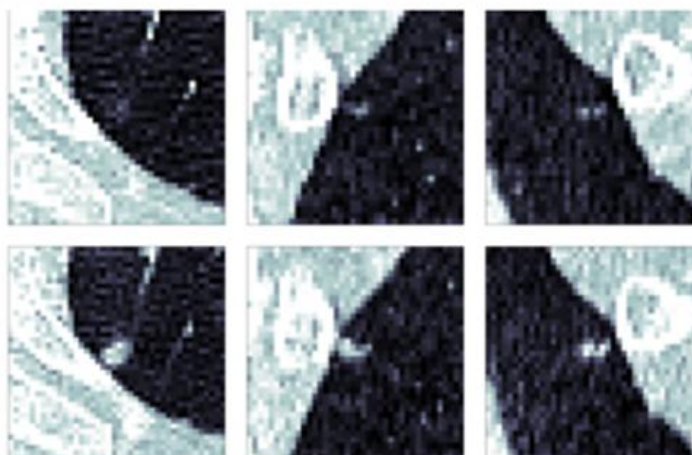
Lungenrundherd #1, linker, oberer
 Lungenlappen (Schicht = 57)



axial

koronar

sagittal



Durchmesser [mm]	8 – 10
Volumen [mm ³]	100 – 250
Veränderung	wachsend (VDT<400 Tage)

Das Visualisierungskonzept wurde mit Thorax Radiologen initialisiert und stetig durch Feedback unsere Kunden verbessert. Weiteres Feedback muss eingeholt und weitere Ideen werden in der verbleibenden Projekt entwickelt, um die Bedürfnisse der Nutzer optimal zu unterstützen. Zwei Beispielbefunde je für eine Baseline und für eine Follow-up Untersuchung, die die Gestaltung des Befundberichts mit den gewählten und final festgelegten Visualisierungselementen darstellt, wurde dem Anhang beigefügt.

AP5.2 Analyse von radiologischen Befunden aus der Versorgung und Identifizierung von Textbausteinen zur strukturierten Befundung

Der Adressat der automatisch zu erstellenden strukturierten Befundberichte ist der Radiologe. Es wurden daher radiologische Befunde aus der Versorgung von zwei unterschiedlichen Zentren, der Radiologie in Erbach und vom assoziierten Partner, der Conradia in Hamburg, eingesammelt und von JDX analysiert. Zusätzlich wurde die Norm DIN 25300-1 mit dem Titel „Prozesse in der Radiologie – Teil 1: Befundung eines bildgebenden oder bildgestützten Verfahrens“ von Mai 2018 hinzugezogen, um die Anforderungen an die Form des Berichts zu definieren.



Radiologische Musterpraxis
Musterstraße 1
12345 Musterstadt

jung diagnostics GmbH
Health Innovation Port (Philips Campus)
Röntgenstr. 24
D-22335 Hamburg

Für Rückfragen:
Tel: 040 / 328 907 93 - 2
service@jung-diagnostics.de

Hamburg, 27.11.2024

CT-Thorax-Untersuchung am 14.11.2024 von Max Mustermann (geb. 01.01.1957)

Sehr geehrte Kolleginnen und Kollegen,

vielen Dank für die Beauftragung einer KI-gestützten, biometrischen Auswertung für o.g. Patienten/in.

Indikation: Die rechtfertigende Indikation wurde durch den/die diensthabenden Arzt/Ärztin bestätigt.

Fragestellung: Lungenkrebscreening/pulmonale Rundherde

Technik: Die Auswertung erfolgte mit Biometrica® auf Basis folgender Bildserien (Siemens Healthineers, SOMATOM go.All): THX-LowDose 1,00 BI56 S3, Schichtdicke: 1.0 mm, Convolution Kernel: BI56f3.

Befund: Die KI-gestützte, biometrische Analyse hat Folgendes ergeben:
Zwei Lungenrundherde (größer als 3 mm).

Mit freundlichen Grüßen,

jung diagnostics GmbH

Abbildung 6. Die erste Seite des strukturierten Befundberichts, der die vom Algorithmus detektierten und quantifizierten Veränderungen im Lungenparenchym zusammenfasst. Die Seite ist so gestaltet, dass sie als Vorbefund gelten kann und vom dem befundenen Radiologen nur korrigiert und ergänzt

werden muss. Dieser Vorbefund kann auch separat über eine HL7-Schnittstelle an das Radiologie-Information-System (RIS) der Radiologie gesendet werden.

Auf Basis dieser Datengrundlage wurden Textbausteine identifiziert, die Ergebnisse des Algorithmus beschreiben. Diese Textbausteine wurden in einer Excel-Tabelle organisiert. In einer Spalte werden die Biomarker-Ergebnisse aufgeführt, in einer weiteren Textspalte die korrespondierenden Textbausteine. Der Algorithmus zur Erstellung eines automatischen Befundes liest die Ergebnisse der automatische Bildanalyse und übersetzt diese dann in die vorgegebenen Textbaustein. Am Schluss wird der Text noch durch eine Nachprozessierung so bereinigt, dass Redundanzen beseitigt und der Gesamtbericht konsistent und leserlich wird. Erste Textbausteine zur Befundung solcher Lungenrundherde wurden erstellt und von Thorax Radiologen bestätigt. Dabei wird das Auswerteschema LungRads 2022 als Vorbild genommen, da dieses Schema für das ein zukünftiges Lungenkrebscreening als Befundungsgrundlage von der deutschen Röntgen Gesellschaft vorgeschlagen wurde. Dieser Vorgang erstreckte sich über einen längeren Zeitraum, weil immer wieder Feedback eingesammelt und Textbausteine feingeschliffen wurden. Die so entworfenen Befunde wurden über einen längeren Zeitraum immer wieder optimiert, indem sie beteiligten Radiologen zur Korrektur vorgelegt wurden. Die vorgeschlagenen Verbesserungen wurden gesammelt, zentral bewertet und eingepflegt. Dieser iterative Prozess wurde drei Mal vollständig durchlaufen.

Die im Anhang 1a und 1b abgelegten Beispielbefunde enthalten auch den vollautomatisch generierten Vorbefund mit dem vom Algorithmus vollautomatisch erstellten strukturierten Befundtext (1. Seite).

AP5.3 Zusammenführung der SPIRABENE-Systemkomponenten in einem SW-Demonstrator

Als erster konkreter Anwendungsfall für die neuen Technologien wurde die Lungenrundherd-Erkennung auf Basis von Niedrig-Dosis CT-Untersuchungen ausgewählt. Für diesen Anwendungsfall wurde ein dedizierter Demonstrator in Python realisiert und in die JDX-proprietäre und kommerziell Bildverarbeitungspipeline (Biometrica Version 1.1) als Forschungsmodul (SPIRABENE-Modul) integriert. Die Integration in Biometrica war notwendig, damit die Lungenrundherd-Erkennung auf Daten aus der Routineversorgung ohne manuellen Aufwand angewendet werden kann. Die Integration in den Auswerteworkflow ermöglicht den Einsatz einer Teleradiologie-Plattform (ReportX; <https://befund.jung-radiology.de/login>), die für die Bearbeitung von AP6.2 (Verbesserung klinischer Workflow) und AP6.3 (Versorgungsnahe Pilotstudie) essentiell ist.

Der Demonstrator weist folgende Verarbeitungsschritte / Module auf:

Die DICOM-Rohdaten eines Kunden/Pilotkunden werden an JDx geschickt, wo sie anonymisiert und angelegt werden. Passen die Daten zu einer zuvor festgelegten Qualified-Sequenz, die die Akquisitionsparameter beschreibt, wird automatisch ein CT-Job angelegt. Dieser startet dann die Python-Pipeline zur Auswertung und Visualisierung.

Der Prozess besteht im Wesentlichen aus 6 Schritten.

1. Abfrage der notwendigen Informationen aus der Datenbank. Es wird ein Objekt Jobelement einer übergeordneten ThoraxCT-Klasse erzeugt, welches bei der Initialisierung die Notwendigen DB abfragen durchführt. Diese Informationen werden bei der Initialisierung der LungNoduleDetection-Klasse übergeben, welche die weiteren Schritte als Methoden steuert.
2. Preprocessing: Für den Job wird ein Workspace angelegt und die zugehörigen DICOM-Daten kopiert. Die Dicom Dateien werden dann in das komprimierte nifti Format überführt.
3. Die im nifti Format gespeicherten Bilddaten werden zunächst auf einen Atlas registriert, um grob die Region der Lunge im Bild zu erkennen und Regionen außerhalb der Lunge (Hals, Abdomen) von der Analyse auszuschließen. Bei der Registrierung auf den Atlas werden grobe Segmentierungen von Anatomien (u.a. der Knochen) erstellt, die im späteren Verlauf verwendet werden.
4. Auf dem groben Lungenausschnitt werden die Lungenlappen segmentiert. Das Bild wird weiter auf diesen Ausschnitt zugeschnitten.
5. Potentielle Lungen-Nodule werden von einem sensitiven Netzwerk segmentiert und dann mittels Klassifikation in falsch positive und tatsächliche Nodule unterteilt. Die als falsch positiv klassifizierten potentiellen Nodule werden aussortiert. Für die restlichen Nodule werden Volumen und Durchmesser gemessen. Außerdem wird hinterlegt, in welchem Lungenlappen und in welcher CT Schicht sie sich befinden.
6. Visualisierung: Mit Hilfe der Maske der Knochen (Schritt 3), der Lungenlappen (Schritt 4) und der Lungen-Nodule (Schritt 5) wird ein Übersichtsplot angefertigt, bei dem, ähnlich wie bei einem Röntgenthorax Bild, die Frontalansicht gewählt wird, um alle Nodule auf einen Blick darzustellen. Für jedes Nodul werden vier Ansichten erstellt: eine Übersicht der axialen Schicht, in der das Nodul am besten zu erkennen ist und in dem das Nodul durch eine rote Box markiert wurde, und drei kleine Ausschnitte des Noduls, in axialer, koronaler und sagittaler Ansicht.
6. Erstellung des pdf-Berichts (AP5.1 & 5.2).

Der erzeugte Bericht kann anschließend vollautomatisch zurückgeliefert werden. Prototyp (als Forschungsmodul von Biometrica, der kommerziellen KI-Engine von jung diagnostics GmbH) ist in mehreren Pilotzentren im Einsatz.

Nach Fertigstellung des Demonstrators wurde eine Gebrauchsanweisung für Pilotkunden erstellt, die Auskunft über Zweckbestimmung der Software, Patientenzielgruppe, klinischen Nutzen, vorgesehene Anwender, Nutzerumgebung und Anforderungen an die Bilddaten gibt. Auch findet sich darin eine komplette Beschreibung des von der Software automatisch erstellen Befundberichts. Die komplette Gebrauchsanweisung liegt als Anhang 2 diesem Schlussbericht bei.

AP6.1 Retrospektive Evaluierung

Eine erste retrospektive Validierungsstudie konnte durchgeführt werden. Zunächst wurde eine 3-fold cross-validation Analyse auf den Trainingsdaten, dem LIDC Datensatz (<https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254>) durchgeführt. Dazu wurde der Datensatz in drei gleich große Teile unterteilt und je ein Netzwerk auf zwei Teilen trainiert und auf dem verbleibenden getestet. Typischerweise ist die Performance des Netzwerkes auf Bildern des Trainingsdatensatzes höher als auf Bildern, mit denen das Netz nicht trainiert worden ist. Durch die Cross-Validation Methode kann die Performance für jeden Scan bestimmt werden, ohne dass dieser für das Training verwendet wurde. Die LIDC Daten wurden von 4 Ärzten unabhängig voneinander annotiert, wobei Nodule mit einem Durchmesser von mindestens 3mm segmentiert und die von einem Durchmesser von weniger als 3mm mit einer Landmarke markiert worden sind. In unserer Analyse ergab sich, dass auch einige mit Landmarken annotierten Nodule einen Durchmesser von bis zu 6mm haben. Auf Grund dieser Inkonsistenz haben wir die mit Landmarken markierten Nodule nochmal mit einer automatischen Lösung segmentiert, und den Schwellwert zur Unterscheidung von großen und kleinen Nodulen auf 6mm gesetzt. Bei der Berechnung der inter-reader Ergebnisse wurden die Größenschwellwerte nur einseitig angewendet, d.h. ein 6mm großes Nodul von Reader 1 galt auch als annotiert, wenn die Annotation von Reader 2 einen Durchmesser von < 6mm ergab.

Die Ergebnisse der Cross-Validation im Einzelnen (Durchmesser $d > 6\text{mm}$):

	Sens	PPV	False-Positive-Rate
Inter-reader	81,8	81,8	0,22
KI	84,8	60,0	0,71

Durchmesser $3\text{mm} < d < 6\text{mm}$:

	Sens	PPV	False-Positive-Rate
Inter-reader	71,5	71,5	0.72
KI	36,7	60	0.66

Im Vergleich zur aktuell besten kommerziellen Lösung Veye, konnten wir auf dem LIDC Datensatz eine gleiche Anzahl an falsch positiven Annotationen feststellen (ca 0.6 pro scan) bei einer leicht höheren Anzahl an falsch negativen Annotationen (0.143 vs 0.271). Umgerechnet bedeutet dies, dass Veye einen Lungenrundherd auf jedem siebten, und unsere Methode auf circa jedem vierten Scan übersieht. Die Anzahl der pro Scan korrekt detektierten Rundherden lag bei 1.12 für Veye und 1.0 bei unserer Lösung. Das bedeutet, dass von den 1.6 Rundherden, die unsere Lösung detektiert, circa ein Drittel falsch positiv und zwei Drittel korrekt sind. Durch die Nutzung der Cross-Validation-Strategie konnten drei Netzwerke trainiert werden, allerdings konnte nur ein Netzwerk pro Scan zur Auswertung genutzt werden. Für neue unabhängige Daten können alle drei Netzwerke ausgewertet werden und ein Mehrheitsvotum angesetzt werden. Das bedeutet, dass ein Rundherd nur als detektiert gilt, falls mindestens zwei von drei Netzwerken den Fund bestätigen. Dies kann die Anzahl an falsch positiven Funden drastisch reduzieren. Zur Validierung unter realen Bedingungen wurden 78 Scans aus der Versorgung annotiert.

AP6.2 Verbesserung klinischer Workflow (Ziel ist Einsparung von Befundungszeit)

Zur Vorbereitung eines Versuchs zur Überprüfung der Einsparung von Befundungszeit wurde eine Gebrauchsanweisung für die Nutzer ausgearbeitet (siehe auch AP5.3). Da die Software zur computerassistierten Detektion primär für den Einsatz im Rahmen der staatlich organisierten Lungenkrebs-Früherkennung vorgesehen ist, wurde die Gebrauchsanweisung in einem zweiten Schritt auf die gesetzlichen Vorgaben ausgerichtet. Diese geben einen konkreten Workflow vor. Die Verordnung legt fest, dass die Computertomographieaufnahme zunächst ohne und anschließend unter Nutzung einer Software zur computerassistierten Detektion befundet werden muss. Dieser gemäß der Verordnung vorgegebene Workflow wurde in zwei Pilotzentren implementiert und auf seine Praktikabilität getestet. Die Tests und das Feedback der beteiligten Radiologen sind in die Optimierung der Darstellung der Ergebnisse eingeflossen (AP5). Der Ergebnis der Tests zeigt, dass die Software den gesetzlich vorgegebenen Workflow optimal unterstützt.

Aufgrund der gesetzlichen Anforderungen wird keine Einsparung an Befundungszeit im Rahmen des staatlich organisierten Lungenkrebscreening-Programms erwartet. Für sekundäre Anwendungen außerhalb des Programms muss die Software ggfs noch angepasst werden.

AP6.3 Versorgungsnahe Pilotstudie (multizentrisch)

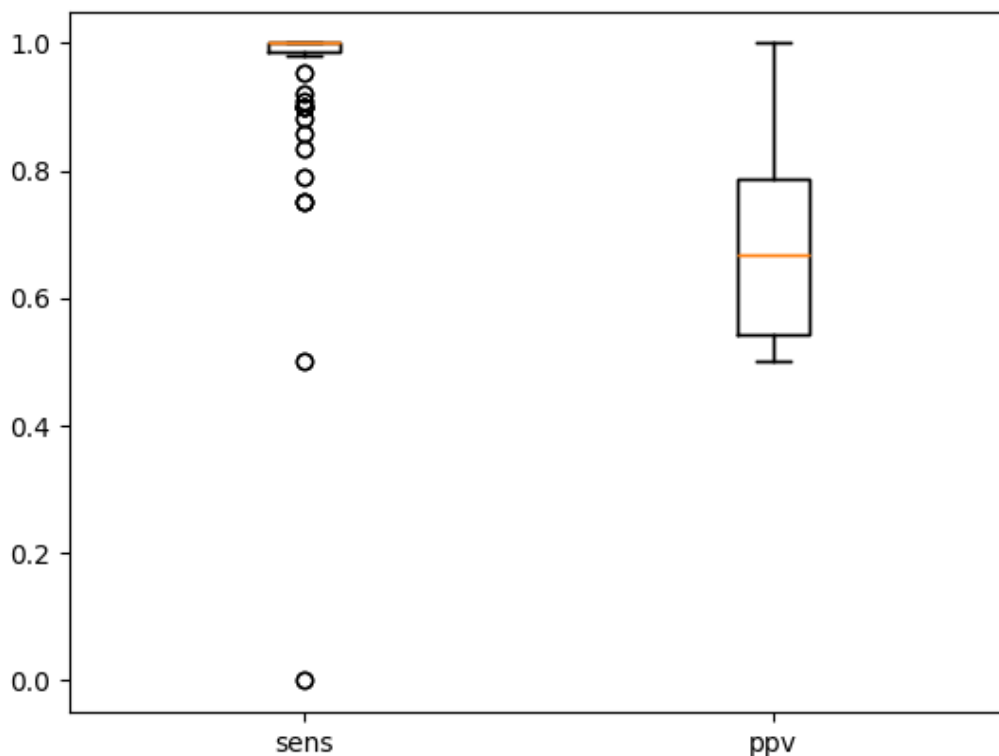
Für die Validierung in der Versorgung wurden n=78 Scans multizentrisch von verschiedenen Kunden gesammelt und annotiert. Bei der Annotation sollten insbesondere Nodule ab 34 mm³ annotiert werden, da diese nach dem LungRADS2022 Auswerteschema als potentiell maligne betrachtet werden. Diese Annotation wurde verwendet um die Sensitivität des Algorithmus zu bestim-

men. Da der Algorithmus auch kleinste Noduli erkennt die mit dem bloßen Auge schnell übersehen werden, wurde das PPV bestimmt indem die Funde der KI visuell begutachtet wurden und in korrekte und falsch positive Funde klassifiziert wurden.

Volumen $v > 34\text{mm}^3$

	Sens	PPV	False-positive rate
KI	94,5	68,8	3,718
Max. Ziel	95,0	60,0	

Für die versorgungsnahe Pilotstudie haben wir zwei Pilotzentren ausgewählt (Radiologie Heinsberg und die radprax-Gruppe).



Die Lungen-CT-Daten aus der Routineversorgung werden von den Biometrica verarbeitet und speziell vom SPIRABENE-Modul ausgewertet. Die automatisch generierten Annotationen als auch die automatisch generierten Befundberichte werden in einer Datenbank abgelegt. Die erzielten Ergebnisse für Sensitivität (sens) und PPV erfüllen oder übertreffen die max. Zielsetzung.

2.2 Wichtigste Positionen des zahlenmäßigen Nachweises

Die für die Durchführung dieses Projekts notwendigen Ausgaben beliefen sich auf insgesamt EUR 615.229,43. Der Eigenanteil der jung diagnostics GmbH betrug EUR 276.979,43.

Insgesamt wurden Fördermittel in Höhe von EUR 338.250,00 für Personalausgaben angefordert und ausgezahlt. Reisekosten wurden nicht abgerechnet. Dennoch wurde das Budget voll ausge-

schöpft, da die zur Verfügung stehenden Reisemittel in Höhe von EUR 3.000,00 in die Finanzierung von Personal geflossen sind (Änderungsbescheid vom 14.02.2025). Diese zusätzlichen Mittel kamen dem Arbeitspaket 6.1 (retrospektive Evaluierung) zugute. Der Bedarf ergab sich dadurch, dass die im Rahmen von AP6.1 (retrospektive Evaluierung) benötigten 150 Patientensätze manuell zwar schon annotiert worden sind, aber die darauf basierende retrospektive Evaluierung der neuen Technologie noch keine statistisch sicheren Ergebnisse produzieren konnte (aufgrund einer höheren Variabilität der Daten, die wir so nicht erwartet haben). Eine neue Abschätzung zeigte, dass weitere 170 Patientendatensätze manuell annotiert werden müssen, damit das Ziel erreicht werden konnte.

2.3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Notwendigkeit

Mit diesem nun erfolgreich abgeschlossenen Förderprojekt konnten wir die bestehenden Technologiebasis, die bislang auf relativ starre Strukturen, wie den Kopf, limitiert war, auf flexible Organe, wie die Lunge, erweitern. Bislang hatte JDX ausschließlich CT und MRT-Anwendungen im Bereich des Kopfes im Portfolio. Mit diesem Förderprojekt wurde eine Technologiebasis für die Detektion und Quantifizierung von kleinsten Veränderungen in hochflexiblen Strukturen, wie des Lungenparenchyms, geschaffen.

Weltweit beobachtet man in den letzten Jahren eine Art „Goldgräberstimmung“ in Bezug auf die Weiterentwicklung von Methoden und Erschließung von Anwendungen der künstlichen Intelligenz (KI). Diese Goldgräberstimmung hat auch die Radiologie erfasst, in der mit Hilfe von neuen neuronalen Netzen, insbesondere DNNs, deutlich technologische Fortschritte erreicht wurden. Diese Fortschritte ziehen Investoren an. Viele neue Start-ups sind weltweit in diesem Bereich in den letzten Jahren entstanden. Direkte Konkurrenten, wie Icometrix (www.icometrix.com) aus Belgien und Mediaire (www.mediaire.ai) aus München haben bislang von Investitionen im niedrigen zweistelligen Millionenbereich profitiert. Damit JDX mithalten kann, muss JDX trotz der schon großen Investitionen und Anstrengungen, die sie in diesem Bereich getätigt hat, seine technologische Basis vertiefen und schnell weiter ausbauen. Im Bereich der Hochtechnologie haben wir im Vergleich zur direkten Konkurrenz einen Vorsprung, den wir uns in den letzten Jahren konsequent erarbeitet haben. Dies wird durch die vergleichsweise große Anzahl an wissenschaftlichen Fachartikeln, die in international anerkannten Fachjournalen (peer-reviewed mit hohem Impaktfaktor) in den letzten Jahren erschienen sind, dokumentiert (<https://www.jung-diagnostics.de/research>). Diesen Vorsprung müssen wir gegen die mit hohen Investitionen ausgestatteten Konkurrenten mindestens halten. Auch mit dem Angang, die Technologie über eine Dienst-

leistung in den Markt einzuführen, die durch eigene Experten (Techniker und Radiologen) qualitätsgesichert wird, besitzen wir ein Alleinstellungsmerkmal. Mit dieser Dienstleistungsplattform haben wir ein Ökosystem aufgebaut, das uns erlaubt, neue technologische Lösungen früh in den Markt einzubringen, dort weiter zu entwickeln und damit in der Patientenversorgung ausreifen zu lassen. Dieses Förderprojekt stellte für JDX einen neuen Meilenstein in der Erweiterung der technologischen Basis von JDX dar. Es hilft entscheidend mit, den technologischen Vorsprung abzusichern.

Angemessenheit

Die Projektmittel wurden sparsam eingesetzt. Reisemittel wurden nicht veranschlagt, da die Treffen mit Kundenbesuchen kombiniert wurden oder Pandemie-bedingt per Videoschleife durchgeführt werden konnten. Angemessen waren auch die Personalmittel, die in dem benötigten Umfang zur Verfügung standen und effizient und sinnvoll eingesetzt wurden. Anzumerken ist die hohe Leistungsbereitschaft der Mitarbeiter des Projekts, die mit viel Energie und Motivation die Aufgaben angegangen sind und immer wieder kreative Lösungen entwickelt haben.

2.4 Nutzen

Primärer Nutzen:

Im Rahmen dieses Förderprojekts wurden grundlegende Bildverarbeitungstechnologien zur automatischen Analyse und Befundung von CT-Untersuchungen der Lunge erforscht und bis zur Anwendungsreife entwickelt. Das wesentliche Ergebnis des Teilvorhabens von JDX ist eine vollautomatisch arbeitende Bildverarbeitungs-Pipeline (SPIRABENE-System) auf Basis von DL-Technologien zur Erkennung kleinster pathologischer Veränderungen im Lungenparenchym. Das SPIRABENE-System besteht aus zwei zentralen technologischen Komponenten: einer Detektion von kleinsten (unspezifischen) pathologischen Veränderungen bei Vorhandensein einer Aufnahme des Patienten und einer Komponente, die bei Vorhandensein einer Voruntersuchung vom selben Patienten die Veränderung zwischen den beiden Untersuchungen quantifizieren kann. Als erster Anwendungsfall wurde das SPIRABENE-System für die Lungenrundherd-Erkennung auf Basis von Niedrig-Dosis CT-Untersuchungen optimiert und validiert. Sie soll nach erfolgter Medizinproduktzulassung in den Markt eingeführt werden und erfüllt vollumfänglich die Anforderungen an eine Software zur computerassistierten Detektion, wie sie in der Verordnung über die Früherkennung von Lungenkrebs bei rauchenden Personen (Lungenkrebs-Früherkennungs-Verordnung (LuKrFrühErKV), veröffentlicht am 17. Mai 2024 im Bundesgesetzblatt, definiert wurden. Die Lungenkrebs-Früherkennung soll ab Frühjahr 2026 großflächig in Deutschland ausgerollt werden.

Sekundärer Nutzen:

- Im Rahmen des SPIRABENE-Projekts wurden auch Technologiekomponenten zur Detektion von Anomalien, wie Milchglatrübungen (ground glass opacities; GGOs) im Lungenparenchym, entwickelt, die im Kontext eines opportunistischen Screenings von Lungenfibrosen eingesetzt werden könnten. Ein entsprechendes Konzept wird aktuell mit einer Pharma-Firma erarbeitet, die ein Interesse hat, Patienten mit Lungen-Fibrosen früh im Erkrankungsstadium zu detektieren.
- Aus dem SPIRABENE-Projekt ist die Idee für ein Eurostars-Projekt (ALTRIS - Softwarelösung zur automatischen Erfassung und Analyse des Krankheitsverlaufs von Lungenkrebs und interstitieller Lungenerkrankung in CT-Bildern) entstanden. Die Projektidee konnte erfolgreich mit dem deutschen Partner Fraunhofer MEVIS und internationalen Partnern, der Uni Nijmegen und der niederländischen Firma Thirona, umgesetzt werden (Förderkennzeichen: 01QE2404A; Laufzeit: 01.05.2024 bis 30.04.2027). Bei ALTRIS liegt der Fokus auf der Entwicklung einer vollautomatisch arbeitenden Bildverarbeitungstechnologie zur Größen- und Volumenmessung von Lungen-Tumoren und interstitiellen Lungenerkrankungen auf Basis von zwei 3D-Thorax-CT-Aufnahmen, die zu unterschiedlichen Zeitpunkten vom gleichen Patienten aufgenommen worden sind. Während des bei SPIRABENE um kleine pathologische Veränderungen im Lungenparenchym ging, entwickelt ALTRIS Technologien, um die Vermessung von Veränderungen von gesicherten Lungentumoren zu erfassen. Es soll untersucht werden, inwieweit man eine Beurteilung von Tumorprogression nach den internationalen RECIST-Standards automatisieren kann. Ohne die ersten Erfolge im SPIRABENE-Projekt, das erst die Grundlagen für eine erfolgreiche Registrierung des Lungenparenchyms gelegt hat, wäre ein solches Projekt zu risikoreich gewesen.

2.5 Fortschritt durch andere

Die Firma Aidence, nun DeepHealth, hat mit wissenschaftlich-klinischen Partnern Ergebnisse ihrer Software zur Erkennung von Lungenrundherde veröffentlicht (Murchison et al. 2022).

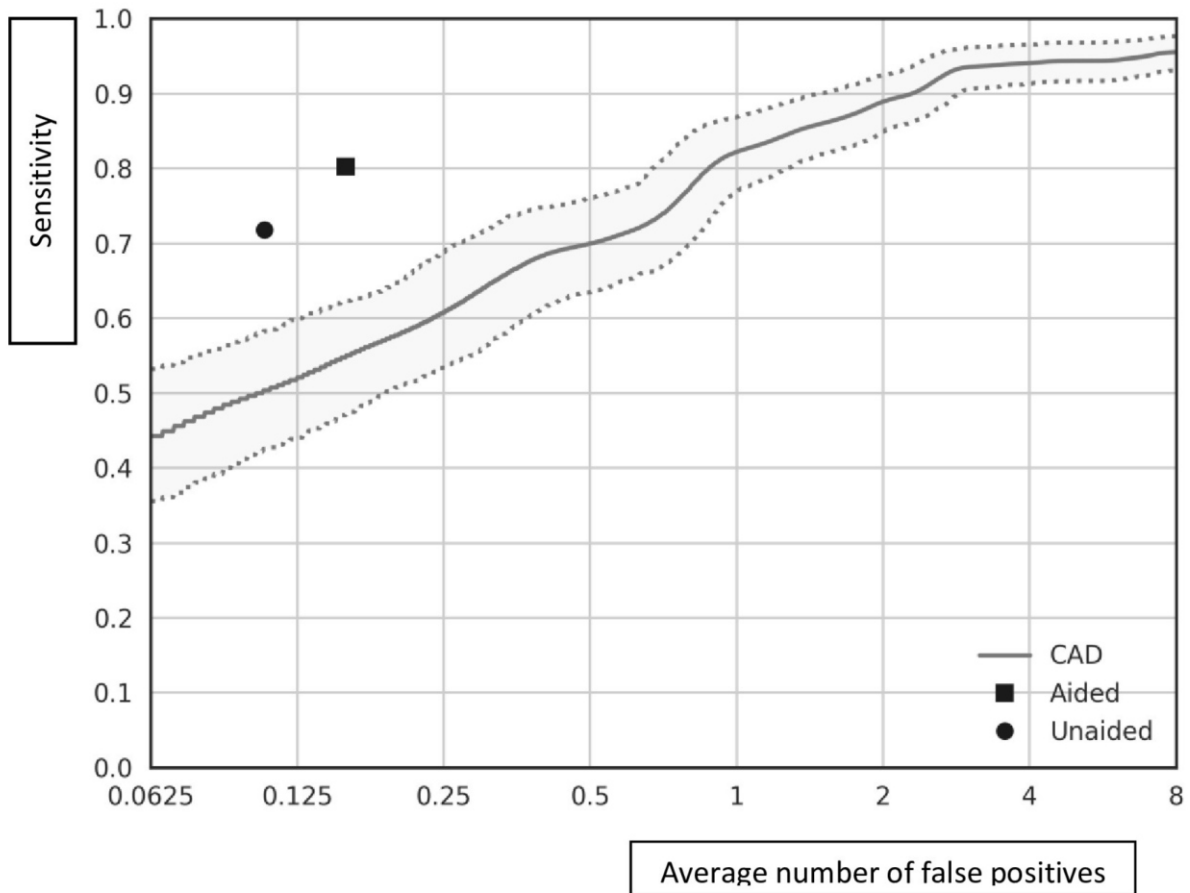


Fig 1. Free-response ROC (FROC) curve. This curve shows the standalone performance of CAD for detecting actionable nodules based on scans from groups 1 and 2. The vertical axis represent the sensitivity and the horizontal axis the average number of false positives per scan. The dashed lines show the upper and lower boundary of 95% confidence interval, bootstrapping of scans with 2000 samples. The circle represents the UNAIDED performance (sensitivity: 71.9% average FP rate 0.11 per scan) and square the AIDED performance (sensitivity: 80.3% average FP rate 0.16 per scan) for detecting actionable nodules.

Murchison JT, Ritchie G, Senyszak D, Nijwening JH, van Veenendaal G, Wakkie J, van Beek EJR. Validation of a deep learning computer aided system for CT based lung nodule detection, classification, and growth rate estimation in a routine clinical population. PLoS One. 2022 May 5;17(5):e0266799.

Software tool	Nodule volume	TP	FP	FN	Sensitivity [%]	PPV [%]
S1	All	2142	316	1203	64	87
	≥ 34 mm ³	1032	210	142	88	83
	≥ 113 mm ³	234	105	48	83	69
S2	All	1526	538	1819	46	74
	≥ 34 mm ³	782	459	392	66	63
	≥ 113 mm ³	202	305	80	72	40

Table 1. Detection performance of both software tools in all nodules, in nodules with volume ≥ 34 mm³ and nodules with volume ≥ 113 mm³. FN, false negative; FP, false positive; PPV, positive predictive value; S1, software tool 1 dataset; S2, software tool 2 dataset; TP, true positive.

Kondrashova R, Klimeš F, Kaireit TF, May K, Barkhausen J, Stiebeler S, Sperl J, Dettmer S, Wacker F, Vogel-Claussen J. Comparison of AI software tools for automated detection, quantification and

categorization of pulmonary nodules in the HANSE LCS trial. Sci Rep. 2024 Nov 13;14(1):27809. doi: 10.1038/s41598-024-78568-z. PMID: 39537690; PMCID: PMC11561337.

2.6 Erfolgte oder geplante Veröffentlichungen

Erfolgte Veröffentlichung

Buddenkotte et al. 2024 „CTARR: A fast and robust method for identifying anatomical regions on CT images via atlas registration“ Journal of Maschine Learning for Biomedical Imaging. 022 vol. 2, pp. 2067-2088

Geplante Veröffentlichung

Die Ergebnisse der retrospektiven Evaluierung werden aktuell zu einem Fachaufsatz zusammengefasst und sollen gegen Ende des Jahres gemeinsam mit den Projektpartner in einer internationalen Fachzeitschrift veröffentlicht werden.

3. ANHANG

1. Lungen-CT-Beispielberichte
2. Gebrauchsanweisung für die computerassistierte Detektion von Lungenrundherden gemäß den gesetzlichen Vorgaben für die Lungenkrebs-Früherkennung.

Radiologische Musterpraxis
Musterstraße 1
12345 Musterstadt

jung diagnostics GmbH
Health Innovation Port (Philips Campus)
Röntgenstr. 24
D-22335 Hamburg

Für Rückfragen:
Tel: 040 / 328 907 93 - 2
service@jung-diagnostics.de

Hamburg, 27.11.2024

CT-Thorax-Untersuchung am 14.11.2024 von Mustermann, Max (geb. 01.01.1957)

Sehr geehrte Kolleginnen und Kollegen,

vielen Dank für die Beauftragung einer KI-gestützten, biometrischen Auswertung für o.g. Patienten/in.

Indikation: Die rechtfertigende Indikation wurde durch den/die diensthabenden Arzt/Ärztin bestätigt.

Fragestellung: Lungenkrebscreening/pulmonale Rundherde; Vergleich mit der Voruntersuchung vom 12.06.2024.

Technik: Die Auswertung erfolgte mit Biometrica® auf Basis folgender Bildserien (Siemens Healthineers, SOMATOM go.All): THX-LowDose 1,00 BI56 S3, Schichtdicke: 1.0 mm, Convolution Kernel: BI56f3.

Befund: Die KI-gestützte, biometrische Analyse hat Folgendes ergeben:
Zwei Lungenrundherde (größer als 3 mm), davon ist einer stabil (VDT > 600 Tage) und einer wachsend (VDT < 600 Tage). Es wurden keine neuen oder abnehmenden Rundherde gefunden.

Mit freundlichen Grüßen,

jung diagnostics GmbH

Biometrica-Report

Name: Mustermann, Max

G-Datum: 01.01.1957

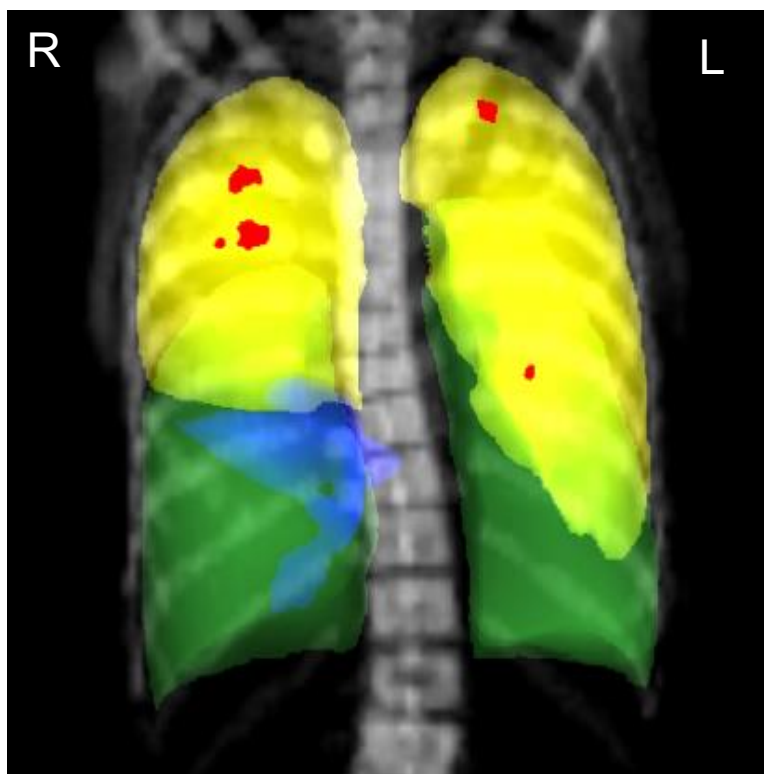
U-Datum: 14.11.2024

Zusammenfassung

In der Aufnahme vom 14.11.2024 wurden zwei Lungenrundherde größer als 3 mm (Minestdurchmesser) detektiert.

Lungenlappen	Rechts	Links
Oberlappen	1	1
Mittellappen	0	NA
Unterbappen	0	0

Durchmesser [mm]	Anzahl = 2
> 30	0
15 – 30	0
10 – 15	1
8 – 10	1
6 – 8	0
3 – 6	0



Im Vergleich zur Voruntersuchung vom 12.06.2024 ergibt sich folgende Veränderungen

Neu	Wachsend (VDT < 600 Tage)	Stabil (VDT > 600 Tage)	Abnehmend
0	0	1	1

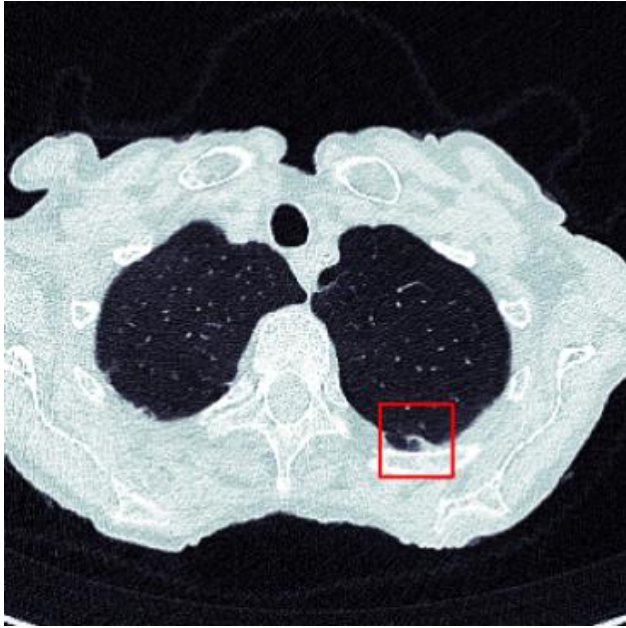
Biometrica-Report

Name: Mustermann, Max

G-Datum: 01.01.1957

U-Datum: 14.11.2024

Lungenrundherd #1, linker, oberer Lungenlappen (Schicht = 57)

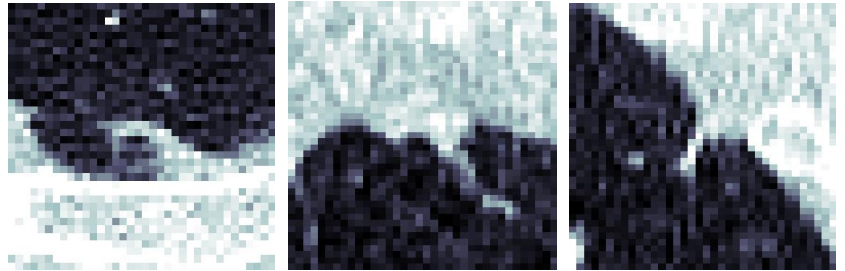


Voruntersuchung vom 12.06.2024

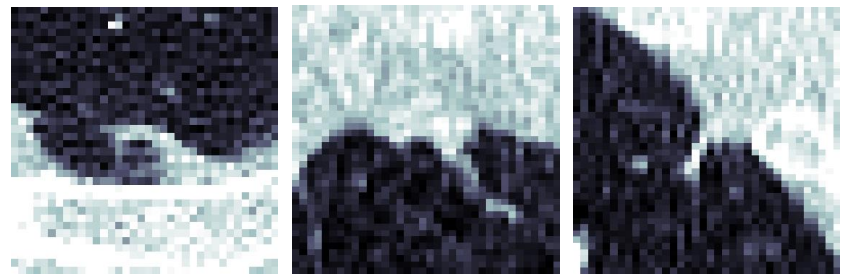
axial

koronar

sagittal

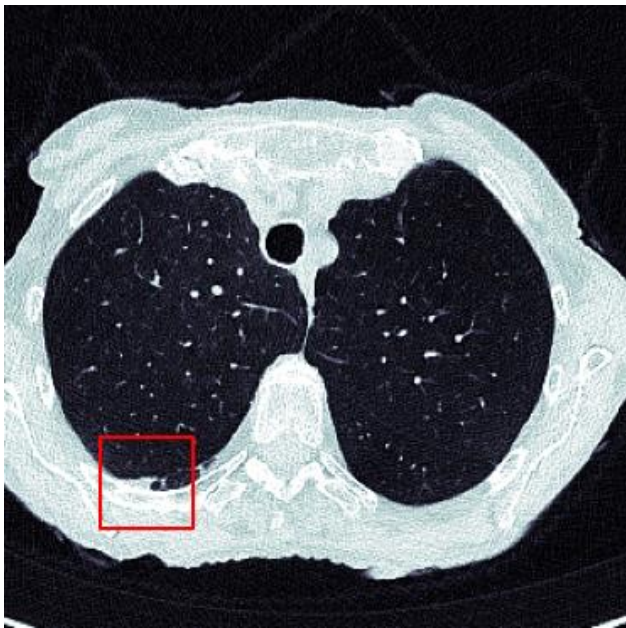


Aktuelle Untersuchung vom 14.11.2024



Durchmesser [mm]	8 – 10
Volumen [mm³]	100 – 250
Veränderung	Neu / wachsend (VDT<600 Tage) / stabil (VDT>600 Tage) / abnehmend

Lungenrundherd #2, rechter, oberer Lungenlappen (Schicht = 98)

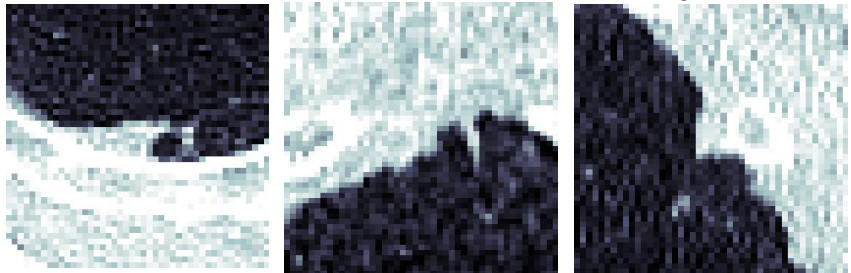


Voruntersuchung vom 12.06.2024

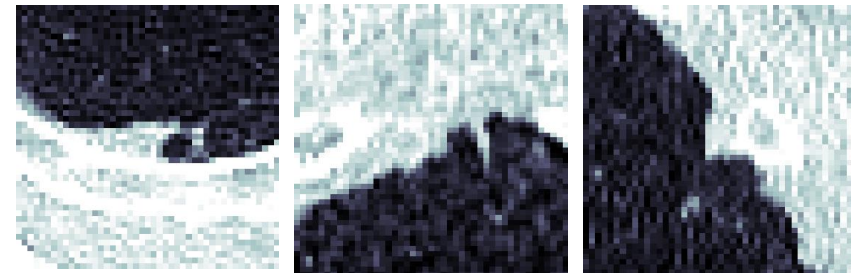
axial

koronar

sagittal



Aktuelle Untersuchung vom 14.11.2024



Durchmesser [mm]	10 – 15
Volumen [mm³]	250 – 500
Veränderung	Neu / wachsend (VDT<600 Tage) / stabil (VDT>600 Tage) / abnehmend

Radiologische Musterpraxis
Musterstraße 1
12345 Musterstadt

jung diagnostics GmbH
Health Innovation Port (Philips Campus)
Röntgenstr. 24
D-22335 Hamburg

Für Rückfragen:
Tel: 040 / 328 907 93 - 2
service@jung-diagnostics.de

Hamburg, 27.11.2024

CT-Thorax-Untersuchung am 14.11.2024 von Max Mustermann (geb. 01.01.1957)

Sehr geehrte Kolleginnen und Kollegen,

vielen Dank für die Beauftragung einer KI-gestützten, biometrischen Auswertung für o.g. Patienten/in.

Indikation: Die rechtfertigende Indikation wurde durch den/die diensthabenden Arzt/Ärztin bestätigt.

Fragestellung: Lungenkrebscreening/pulmonale Rundherde

Technik: Die Auswertung erfolgte mit Biometrica® auf Basis folgender Bildserien (Siemens Healthineers, SOMATOM go.All): THX-LowDose 1,00 BI56 S3, Schichtdicke: 1.0 mm, Convolution Kernel: BI56f3.

Befund: Die KI-gestützte, biometrische Analyse hat Folgendes ergeben:
Zwei Lungenrundherde (größer als 3 mm).

Mit freundlichen Grüßen,

jung diagnostics GmbH

Biometrica-Report

Name: Mustermann, Max

G-Datum: 01.01.1957

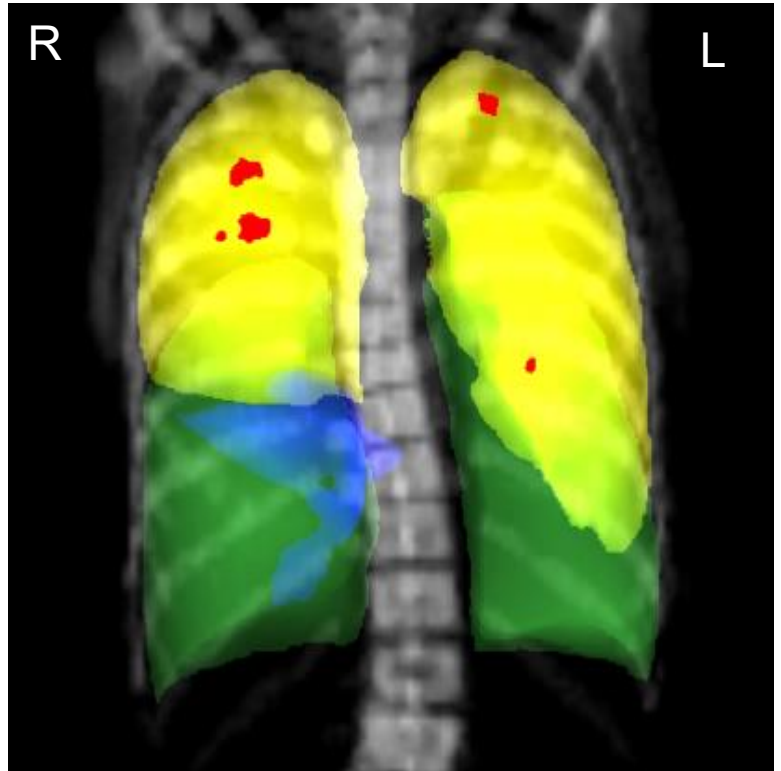
U-Datum: 14.11.2024

Zusammenfassung

In der Aufnahme vom 14.11.2024 wurden zwei Lungenrundherde größer als 3 mm (Minstdurchmesser) detektiert.

Lungenlappen	Rechts	Links
Oberlappen	1	1
Mittellappen	0	NA
Unterbappen	0	0

Durchmesser [mm]	Anzahl = 2
> 30	0
15 – 30	0
10 – 15	1
8 – 10	1
6 – 8	0
3 – 6	0



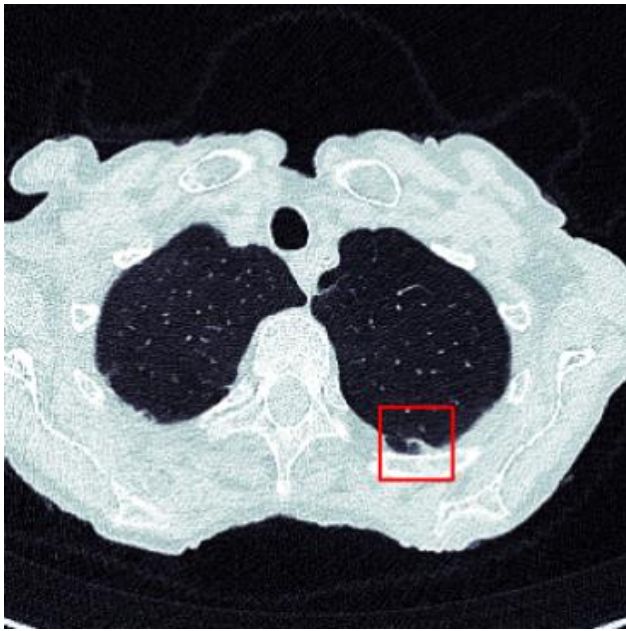
Biometrica-Report

Name: Mustermann, Max

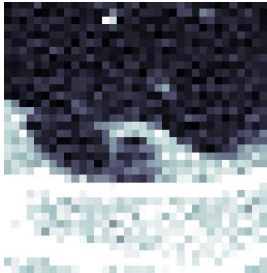
G-Datum: 01.01.1957

U-Datum: 14.11.2024

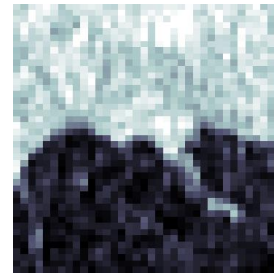
Lungenrundherd #1, linker, oberer
Lungenlappen (Schicht = 57)



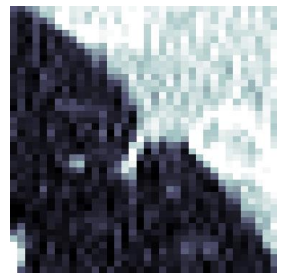
axial



koronar



sagittal



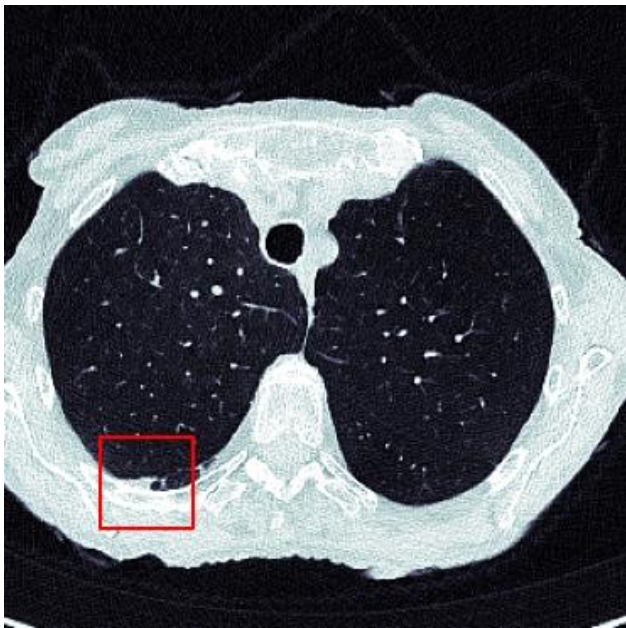
Durchmesser [mm]

8 – 10

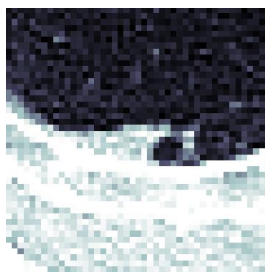
Volumen [mm³]

100 – 250

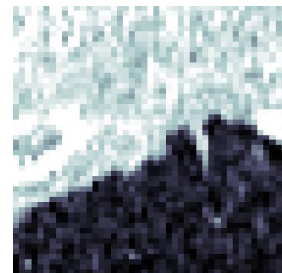
Lungenrundherd #2, rechter, oberer
Lungenlappen (Schicht = 98)



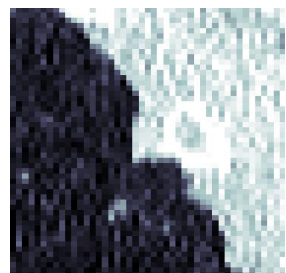
axial



koronar



sagittal



Durchmesser [mm]

10 – 15

Volumen [mm³]

250 – 500

BIOMETRICA (Version 0.1) | Lungen CT

Gebrauchsanweisung

**LESEN SIE DIESE GEBRAUCHSANWEISUNG BEVOR SIE DAS PRODUKT
VERWENDEN.**

Gebrauchsanweisung Version 1.0

Stand Dezember 2024









Inhaltsverzeichnis

LEGENDE (SYMBOLE)	3
WARNUNGEN UND VORSICHTSHINWEISE.....	4
BEGRIFFE UND ABKÜRZUNGEN	5
ALLGEMEINE INFORMATIONEN.....	6
PRODUKTBESCHREIBUNG	7
Zweckbestimmung.....	7
Indikation	8
Patientenzielgruppe.....	8
Vorgesehene Anwender	8
Nutzerumgebung	8
Spezifikation der Bilddaten	9
Beschreibung des Biometrica-Reports.....	11
SICHERHEITS- UND LEISTUNGSANFORDERUNGEN.....	13
Grundlegenden Sicherheits- und Leistungsanforderungen	13
Spezifische Sicherheitsanforderungen.....	14
Spezifische Leistungsanforderungen	14
Klinischer Nutzen.....	14
Fehlerquellen / Limitationen.....	14
Risikobetrachtung.....	15
INSTALLATION, PACS-INTEGRATION UND WARTUNG.....	15
Installation und PACS-Integration	15
Wartung.....	15

LEGENDE (SYMBOLE)

Es gilt, auf folgende Symbole und deren Bedeutung in dieser Gebrauchsanweisung zu achten:

Symbol	Bedeutung
	Herstellungsdatum
	Hersteller
	CE-Zeichen (europ. Konformität)
	Vorsicht
	Medical Device
	Gebrauchsanleitung / Handbuch

WARNUNGEN UND VORSICHTSHINWEISE

Folgende Hinweise geben dem Anwender Auskunft über etwaige Warnungen, Vorsichtshinweise, Kontraindikationen, zu ergreifende Maßnahmen sowie Verwendungsbeschränkungen im Zusammenhang mit dem Produkt.



Die Verwendung von Biometrica setzt die Kenntnis der in der Gebrauchsanweisung gelieferten Informationen voraus.



Biometrica darf nicht außerhalb der Zweckbestimmung verwendet werden (siehe Kapitel „Produktbeschreibung“). Biometrica erstellt keine auf Grundlage der gefundenen Auffälligkeiten basierende Diagnose oder macht einen Diagnose- oder Therapievorschlag. Die alleinige Aufgabe von Biometrica ist, strukturelle Veränderungen auf radiologischen Bilddaten zu detektieren und zu annotieren. Der Anwender darf die Annotationen/Ergebnisse von Biometrica daher nur nach visueller Prüfung und in Kontext mit anderen visuellen Bildmerkmalen bewerten und daraus seinen Befund ableiten und den Befundbrief erstellen. Biometrica ist daher nur ein Hilfsmittel im Sinne der Detektion, kein Interpretationstool. Zur Interpretation bedarf es das Wissen eines für die radiologische Befundung qualifizierten Arztes (Radiologie oder zuweisender Facharzt). Der Hersteller übernimmt keine Verantwortung für eine inkorrekte Interpretation der Ergebnisse.



Die Zweckbestimmung sieht einen Gebrauch der Ergebnisse der Software als eine zusätzliche zur Befundung erstellte Dokumentation vor, die in keiner Weise den Befundungsprozess des Arztes ersetzt. Eine Fehlanwendung liegt vor, wenn der Arzt die Ergebnisse ohne vorherige Prüfung ganz oder teilweise in seinen Befundbericht einbindet.



Die Bildqualität der zugrundeliegenden Lungen CT-Daten nimmt Einfluss auf die Sensitivität von Biometrica. So können z.B. Atmungsartefakte, schlechte Kontrastwerte durch Strahlaufhärtung oder sonstige Artefakte zu einer höheren Anzahl an falsch-positiven Ereignissen führen. Nähere Informationen und Folgen von Fehlerquellen finden sich in Kapitel zu den Leistungsmerkmalen von Biometrica.



Die Sensitivität und Spezifität von Biometrica ist weniger als 100% - strukturelle Veränderungen in der Lunge können unentdeckt bleiben und daher nicht oder unzureichend annotiert sein.



Im Falle von signifikant pathologischen Veränderungen des Lungenparenchyms kann es zu Detektionsfehlern und unerwarteten Abweichungen bzgl. der Performanz kommen, die auch die auch außerhalb der Spezifikation liegen können (siehe Kapitel Leistungsmerkmale)

BEGRIFFE UND ABKÜRZUNGEN

Begriff / Abkürzung	Beschreibung
Biometrica	vereinfachter Name des Medizinprodukts, auf das sich dieses Dokument bezieht; vollständiger Name: Biometrica (Version 1.1) zur Detektion und Dokumentation von strukturellen Auffälligkeiten, z. B. Lungenrundherde auf Lungen CT-Aufnahmen
CNN	Convolutional Neural Network
DICOM	Digital Imaging and Communications in Medicine
DICOM – SC	Digital Imaging and Communications in Medicine - Secondary Capture (ein DICOM Format)
Falsch Positive	Ein falsch positives Ergebnis; in diesem Fall eine Annotation ohne zugrundeliegende strukturelle Auffälligkeit im CT-Bild
Falsch Negative	Ein falsch negatives Ergebnis; in diesem Fall das Ausbleiben einer Annotation trotz struktureller Auffälligkeit im CT-Bild
KI	Künstliche Intelligenz; wird oft als Oberbegriff für Software verwendet, die neuronale Netze und tiefe Lernverfahren (Deep Learning) verwenden.
PACS	Picture Archiving and Communication System
Sensitivität	Qualitätsmerkmal und Maß für die Funktionalität von Biometrica; = $\frac{\text{Richtig Positive}}{\text{Richtig Positive} + \text{Falsch Negative}}$
Spezifität	Qualitätsmerkmal und Maß für die Funktionalität von Biometrica; = $\frac{\text{Richtig Negative}}{\text{Richtig Negative} + \text{Falsch Positive}}$

ALLGEMEINE INFORMATIONEN

Die Gebrauchsanleitung liefert Informationen zur sicheren Anwendung von Biometrica und legt alle Funktionen und die zur Ergebnisinterpretation notwendigen Informationen dar.

Bitte stellen Sie sicher, dieses Dokument gelesen und verstanden zu haben, bevor Sie dieses Medizinprodukt nutzen. Aus Gründen der besseren Lesbarkeit wird nachfolgend auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

Der Anwender dieses Produkts ist allein für alle Fehlfunktionen verantwortlich, die durch unsachgemäße Verwendung und nicht autorisierte Änderungen entstehen. Der Hersteller empfiehlt daher auch die Teilnahme an Produkt-Schulungsmaßnahmen, die vom Hersteller nach Bedarf und auf Anforderung durchgeführt werden.

Der Hersteller behält sich das Recht vor, seine Produkte und die zugehörige Dokumentation regelmäßig zu ändern oder zu verbessern.

Dieses Dokument ist gültig für:

Biometrica Version 1.1 (Erkennung und Detektion von strukturellen Veränderungen z. B. Lungenrundherde auf Lungen CT-Daten)



Herstellungsdatum/Release Datum: Dezember 2024.



Biometrica ist ein Produkt der **jung diagnostics GmbH** (Hersteller). Für zusätzliche Informationen, kontaktieren Sie die jung diagnostics GmbH.

Website: <https://www.jung-diagnostics.de>
Email: info@jung-diagnostics.de
Adresse: Röntgenstraße 24, 22335 Hamburg, Deutschland
Telefon: +49 (0) 40 / 328 907 93 – 9



Die letzte Version dieses Dokuments kann abgerufen werden unter: <https://jung-diagnostics.de/docs>

PRODUKTBESCHREIBUNG

Zweckbestimmung

Biometrica ist eine eigenständige (standalone), auf Methoden der künstlichen Intelligenz (KI) basierende Software, die Lungen-CT-Daten eines Patienten verarbeitet und die strukturelle Veränderungen im Lungenparenchym (insbesondere Lungenrundherde) detektiert, in Größenkategorien einordnet, für den Anwender visuell aufbereitet und in einem Bericht dokumentiert.

Der klinische Nutzen von Biometrica besteht darin, dass Biometrica den Anwender bei der Detektion und Dokumentation von strukturellen Veränderungen (insbesondere Lungenrundherde im Lungenparenchym) unterstützt. Für den Radiologen kann das eine signifikante Zeitersparnis bedeuten, da sich die Dokumentation seines Befunds vereinfacht. Der Zusatznutzen für den Zuweiser und damit indirekt auch für den Patienten liegt darin, dass er eine standardisierte, vollständige Befunddarstellung erhält, die die Vorgaben der aktuellen Leitlinien der medizinischen Fachgesellschaften umsetzt.



Biometrica erstellt keine auf Grundlage der gefundenen Auffälligkeiten basierende Diagnose oder macht einen Diagnose- oder Therapievorschlag.



Biometrica ist eine Detektionshilfe, kein Interpretationstool.



Biometrica zeigt nur strukturelle Veränderungen des Lungenparenchyms an, keine weiteren Abnormalitäten.



Biometrica kann auffällige Regionen markieren, die keine strukturellen Veränderungen des Lungenparenchyms darstellen. Umgekehrt kann es sein, dass strukturelle Veränderungen unerkannt bleiben. Für beide Eventualitäten gilt neben den zugrundeliegenden Spezifitäts- und Sensitivitätsmerkmalen von Biometrica zu berücksichtigen, dass zusätzliche Faktoren (z.B. Bilddatenqualität) Einfluss auf genau diese Maße haben können. Daher muss der Anwender beim Review stets die Interpretationsergebnisse in Kontext mit seinem Expertenwissen setzen (nähere Informationen zu Fehlerquellen im Kapitel „Leistungsmerkmalen“).



Biometrica analysiert nur Lungen CT-Daten, die die richtigen DICOM Metainformationsdaten (z.B. Format, Inhalte des DICOM-Headers etc.) beinhalten (Näheres dazu in Kapitel „Spezifikation der Bilddaten“).



Biometrica ist nur dazu bestimmt, strukturelle Veränderungen im Lungenparenchym (Lungenrundherde ab einem Volumen von 0,01 ml, das einem Mindestdurchmesser von 3 mm entspricht) zu detektieren. Die Performanz nimmt ab, sobald diese Veränderungen einen Durchmesser von kleiner als 3 mm haben. Die Veränderungen werden dann nicht oder nur unzureichend detektiert und annotiert. Über diese Größenschwelle hinaus detektiert der Algorithmus alle Verdichtungen unabhängig von der pathologischen Ursache.

Indikation

Die Anwendung von Biometrica auf radiologische Bilddaten eines Patienten ist indiziert, wenn Patienten Erkrankungen haben oder Erkrankungen vermutet werden, die strukturelle Veränderungen der Lunge hervorrufen können, insbesondere Lungenrundherde im Lungenparenchym.

Biometrica erfüllt vollumfänglich die gesetzlichen Anforderungen an eine Software zur computerassistierten Detektion für die Lungenkrebs-Früherkennung.¹

Eine Kontraindikation liegt nicht vor.

Patientenzielgruppe

Personen, die 18 Jahre oder älter sind, da Biometrica nur für diese Patientengruppe validiert wurde.

Vorgesehene Anwender

Der vorgesehene Anwender ist ein für die radiologische Befundung qualifizierter Arzt (Radiologe) und/oder zuweisender Facharzt. Der Radiologe ist der für die radiologische Untersuchung verantwortliche Arzt, der die Untersuchung durchführt und die Bilddaten gründlich inspiziert und interpretiert. In der Regel werden diese über das PACS, ein für radiologische Bilddaten genutztes Bildarchivierungs- und Kommunikationssystem, befundet. Auffälligkeiten beschreibt der Arzt in einem Arztbrief an den Zuweiser oder zuweisenden Arzt (radiologischer Befund). Der radiologische Bericht enthält ausschließlich Verdachtsdiagnosen. Die für die Behandlung entscheidende Diagnose erstellt nur der Zuweiser oder zuweisende Arzt. Der radiologische Befund ist ein Merkmal in der Gesamtschau aller Merkmale, die in die finale Diagnose des Zuweisers oder zuweisenden Arztes einfließen.

Nutzerumgebung

Die Arbeitsumgebung des Radiologen, der radiologische Befunde erstellt, besteht aus einem PACS-System, in dem die Bilder der Untersuchung gespeichert sind und zusätzlich einem Radiologie-Informationssystem (RIS) oder Krankenhaus-Informationssystem (KIS), in dem die Patientendaten verwaltet und das Befundungsmodul integriert ist. Der Radiologe arbeitet nach einer durch das RIS/KIS vorgegebenen Arbeitsliste die zu befundenen Patienten ab. Ruft er einen Patienten aus der Arbeitsliste zur Befundung auf, öffnen sich automatisch im PACS die Bilddaten der Untersuchung, die dann nach vorgegebenen Protokollen angezeigt werden. Der Radiologe schaut sich alle Bilder an und beginnt den Befund zu diktieren und nutzt dazu in der Regel eine Software zur Spracherkennung. Idealerweise liegt der Biometrica-Report zu diesem Zeitpunkt im PACS-System vor. Der

¹§4 der Verordnung über die Zulässigkeit der Anwendung der Niedrigdosis-Computertomographie zur Früherkennung von Lungenkrebs bei rauchenden Personen (LuKrFrühErkV – vom 17. Mai 2024): https://www.gesetze-im-internet.de/lukrfr_herkv/BJNR0A20A0024.html

Radiologe schaut sich den Biometrica-Report nach dem Verfassen des Befunds an und entscheidet, ob der Biometrica-Bericht seinen Befund unterstützt und die Darstellung eine nützliche Dokumentation seines Befunds darstellt. Entscheidet er sich dazu, fügt er den Biometrica-Bericht seinem Arztbrief bei. Der Arzt kann den Biometrica-Bericht nicht modifizieren. Möchte er dazu einen Kommentar machen, muss er diesen Kommentar in seinem Arztbrief mit aufnehmen. Der Anwender benötigt einen mit einem PACS-System ausgestatteten Computer. Biometrica kann mit allen PACS-Systemen genutzt werden, es erfordert lediglich die Möglichkeit zur Darstellung von Derived DICOMs, optional auch PDFs oder alternativ Secondary Capture Bildformaten. Die Annotationsergebnisse von Biometrica werden in Form von mindestens einer dieser DICOM-Formate zur Verfügung gestellt und sind im PACS einsehbar (nicht für alle Produktmerkmale). Die Verfügbarkeit der Ergebnisse ergibt sich durch ein gesichertes VPN-Gateway, das einmalig vor Ort in das Netzwerk eingebunden wird und über dieses der Versand der anonymisierten DICOM-Daten zu jung diagnostics und der Eingang der Annotationsergebnissen in Form von DICOM-Objekten geregelt wird. Der Datenausgang erfolgt mittels regel-basierten Autorouting oder manuell. Nähere Informationen dazu können Sie dem Kapitel 6 entnehmen. Die Standalone-Software Biometrica wird explizit nicht beim Anwender betrieben, sondern nur auf angemieteten oder eigenen Servern des Herstellers. Der Anwender (Radiologie) hat keinen Zugang zur Software und bedient die Software nicht. Damit besteht zu jeder Zeit die volle Kontrolle des Herstellers über die Ausführung der Auswertung.

Wenn Biometrica im Rahmen der Lungenkrebs-Früherkennung gemäß der Verordnung LuKrFrühErkV vom 17. Mai 2024 genutzt wird, ist sicherzustellen, dass die Computertomographieaufnahme zunächst ohne und anschließend unter Nutzung von Biometrica befundet wird.

Spezifikation der Bilddaten

Für die ordnungsgemäße Anwendung des Produkts muss die Qualität der Bilddaten und die Akquisition der Bilddaten Mindeststandards erfüllen. Denn die Qualität der Auswertung hängt maßgeblich von der Qualität der zugrundeliegenden Bilddaten ab.

Die Spezifikation der Bildakquisition und Bildqualität als Voraussetzung für die Anwendung von Biometrica:

Parameter	Anforderung
Scanmodus	Spirale zur lückenlosen Erfassung in einer Atempause
Scanzeit	≤ 15 Sekunden
Rotationszeit	≤ 0,5 Sekunden
Einzelschichtkollimation	≤ 0,7 Millimeter
Pitch	Zwischen 0,9 und 1,2
Spannungsautomatik	Ja
Dosisautomatik/Röhrenstromautomatik	Verwendung von Protokollen mit an die Körperstatur angepasster Röhrenspannung (z. B. mittels Spannungsautomatik) und

	angepasstem Röhrenstrom (z. B. mittels sektorieller Röhrenstromabsenkung)
Dynamische Kollimation	Hardwarebasierte Methode zur Vermeidung von Overranging bei Detektorabdeckung größer als 4 Zentimeter
Faltungskern	Lunge: kantenbetonter Rekonstruktionsfilter Mediastinum: glättender Rekonstruktionsfilter
Rekonstruktionsverfahren	Modellbasierte iterative Rekonstruktion oder andere Methoden mit mindestens vergleichbarer Dosiseinsparung
Auflösung lateral (xy)	Eine isotrope räumliche Auflösung von 0,8 bis 1 Millimeter für Kontraste ab 150 Hounsfield-Einheiten Die Voxelgröße liegt bei 70 bis 80 % der Auflösung.
Lagerung der zu untersuchenden Person	Rückenlage mit Armen über dem Kopf; zentriert im Isozentrum
Gantrykipfung	0°
Scanrichtung	Caudal-cranial
Spiral-Scan	Durchführung in einer Atempause in tiefer Inspiration
Kontrastmittel	Nein
Scanbereich	Lungenspitze bis dorsaler Recessus
Gesichtsfeld	Gesamte Lunge, ansonsten so gering wie möglich

(teilweise übernommen aus der Lungenkrebs-Früherkennungs-Verordnung – LuKrFrühErkV)

Voraussetzung für die Durchführung der Biometrica-Auswertung ist, dass einmalig zu Beginn das Akquisitionsprotokoll eingestellt und durch die jung diagnostics GmbH freigegeben wurde. Mit dem ersten Datensatz wird die Bilddatenqualität durch die jung diagnostics GmbH manuell überprüft und das Akquisitionsprotokoll hinsichtlich der wichtigsten Parameter im System hinterlegt und technisch so eingestellt, dass eine automatische Auswertung durch Biometrica für alle weiteren Fälle mit dem freigegebenen Protokoll erfolgen kann.

Geplante Änderungen am CT-Protokoll (z.B. Parameter-Einstellungen), an der Hardware oder an der Software (z.B. Updates) müssen unverzüglich an die jung diagnostics GmbH gemeldet werden. Durch Änderungen wird i.d.R. die bestehende Freigabe zur Datenauswertung ungültig und es muss eine neue Qualifizierung durchgeführt werden. Bei Änderungen, die ohne Mitteilung erfolgen und die somit Abweichungen zum voreingestellten Akquisitionsprotokoll darstellen, findet keine automatische Auswertung durch Biometrica statt.

Trotz Einhaltung der o.g. Voraussetzungen können im Routinebetrieb Qualitätsschwankungen in der Bildqualität (Bewegungsartefakte etc.) auftreten. Diese Qualitätsschwankungen können von den Algorithmen trotz der enormen Bandbreite an Trainingsdaten (sog. Grundwahrheit), die auf Basis von Routinedaten erstellt wurden, nicht immer vollständig kompensiert werden und daher zu falschen Ergebnissen (falsch-positive und falsch-negative Ergebnisse) führen.

Hinweise:



Bei anderen Bilddaten mit anderen bzw. nicht den oben definierten Voraussetzungen findet keine automatische Auswertung statt und es wird kein Biometrica Bericht erzeugt.



Weichen im Betrieb die Akquisitionsparameter der eingesendeten Daten von denen des zu Beginn qualifizierten/ freigegebenen Akquisitionsprotokolls um mehr als 10% ab, kann aus Sicherheitsgründen keine vollautomatische Auswertung erfolgen. Dies gilt auch für Daten von Verlaufsuntersuchungen, die alle mit demselben qualifizierten Akquisitionsprotokoll akquiriert werden müssen



Bei niedriger Bildqualität aufgrund von Bildrauschen oder Artefakten wird i.d.R. trotzdem eine automatische Auswertung generiert. Daher gilt es in diesen Fällen besonderes genau auf inkorrekte Ergebnisse zu achten.

Beschreibung des Biometrica-Reports

Der Biometrica Report kann in den Formaten DICOM-PDF oder DICOM-Structured Report dem Anwender optional pro Studie/Fall im PACS zur Verfügung stehen (am Beginn der Kooperation mit der jung diagnostics GmbH festzulegen).

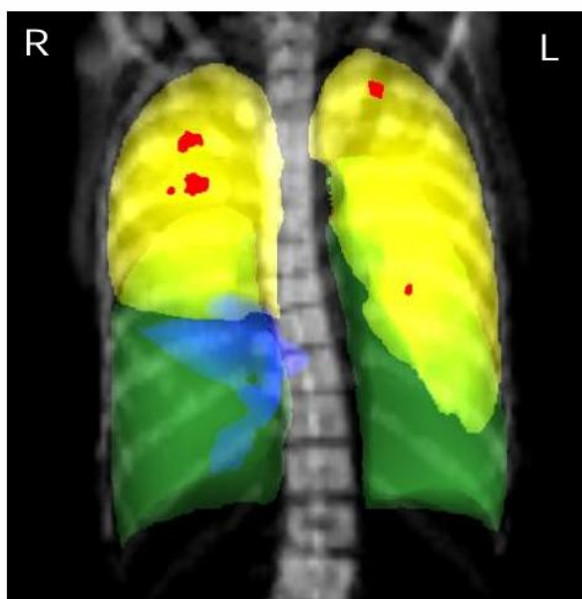
Der Report liefert eine Zusammenfassung der Ergebnisse, u.a. in Textform lesbar sowie in tabellarischer Listung. Pro Annotation/ Detektion eines Lungenrundherds wird ein Bildausschnitt der entsprechenden Region gezeigt.

Zusammenfassung

In der Aufnahme vom 14.11.2024 wurden zwei Lungenrundherde größer als 3 mm (Mindestdurchmesser) detektiert.

Lungenlappen	Rechts	Links
Oberlappen	1	1
Mittellappen	0	NA
Unterbappen	0	0

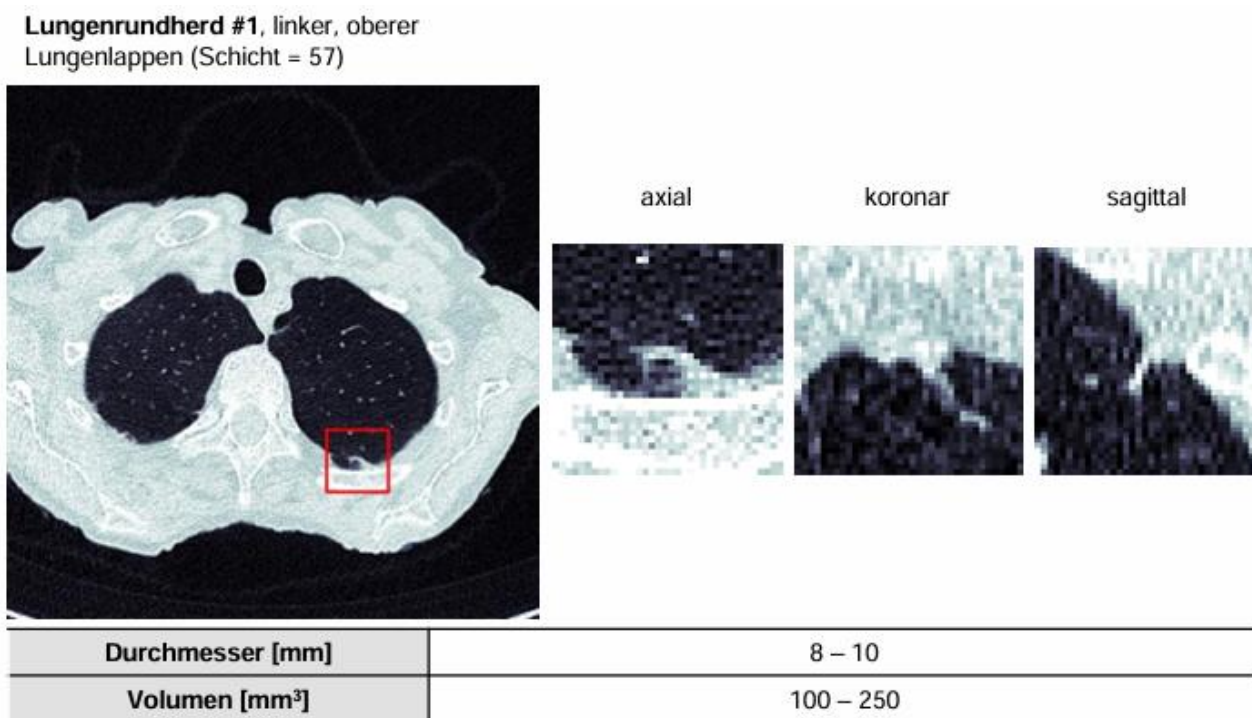
Durchmesser [mm]	Anzahl = 2
> 30	0
15 – 30	0
10 – 15	1
8 – 10	1
6 – 8	0
3 – 6	0



Die gefundenen Lungenrundherde werden lokalisiert (z. B. Oberlappen links) und in Größenklassen (größer > 30 mm, 15 – 30 mm, 10 – 15 mm, 8 – 10 mm, 6 – 8 mm und 3 – 6 mm) eingeteilt.

Man unterscheidet zwei Berichts-Varianten, abhängig davon, ob eine Voraufnahme der untersuchten Person zur Verfügung steht.

Visualisierung eines detektierten Lungenrundherds ohne das Vorhandensein einer Voraufnahme:



Wenn eine Voraufnahme vorhanden ist, wird ein direkter Vergleich der Voraufnahmen mit der aktuellen Aufnahme durchgeführt.

Ist ein Lungenrundherd, der in der Voraufnahme vorhanden ist, auch in der aktuellen Aufnahme persistent, dann wird die Veränderung bestimmt.

Die Veränderung wird durch die Berechnung der Volumenverdopplungszeit (VDT) abgebildet. Diese Größe bestimmt, wie sehr sich das Volumen eines Lungenrundherds zwischen dem Zeitpunkt der Voraufnahme und dem Zeitpunkt der aktuellen Aufnahme verändert hat. Konkret gibt sie die Anzahl der Tagen an, bis sich das Volumen des Lungenrundherds verdoppelt. Zur Berechnung werden die Volumina zum ersten und zweiten Aufnahmezeitpunkt (v_0, v_1) benötigt, sowie die Anzahl der Tagen, die zwischen dem ersten und zweiten Scan vergangen sind ($t_1 - t_0$). Die VDT berechnet sich dann mit folgender Formel:

$$VDT = \frac{t_1 - t_0}{\log_2(v_1) - \log_2(v_0)}$$

Im Fall eines schrumpfenden Lungenrundherds, ist $VDT < 0$ und beschreibt die (negative) Anzahl an Tagen, in der sich das Volumen des Lungenrundherds halbiert.

Lungenrundherde mit einem positiven VDT Wert von unter 400 Tagen werden als schnell wachsend klassifiziert und haben ein hohes Krebsrisiko. Befindet sich der Werte zwischen 0 Tagen und -400 Tagen, ist der Lungenrundherd schrumpfend und hat kein Krebsrisiko. Alle anderen werden als stabil klassifiziert und haben ein geringes Krebsrisiko.

Zusammenfassend wird folgende Tabelle im Bericht ausgegeben:

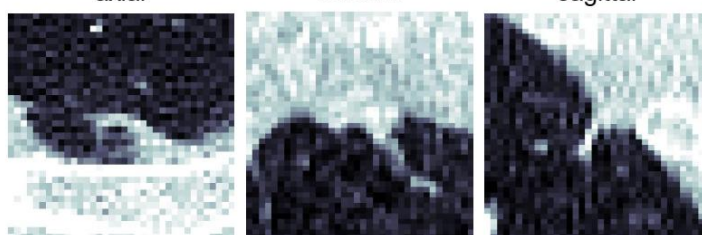
Neu	Wachsend (VDT < 600 Tage)	Stabil (VDT > 600 Tage)	Abnehmend
0	0	1	1

Zusätzlich wird jeder persistente Lungenrundherd wie folgt visualisiert:

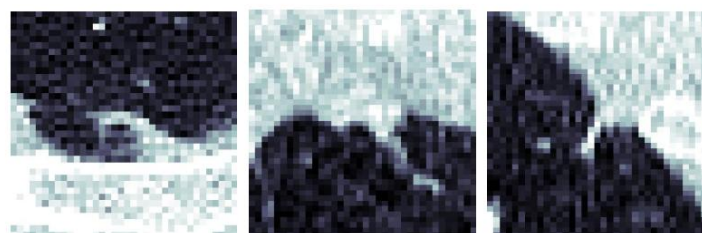
Lungenrundherd #1, linker, oberer Lungenlappen (Schicht = 57)

Voruntersuchung vom 12.06.2024

axial koronar sagittal



Aktuelle Untersuchung vom 14.11.2024



Durchmesser [mm]	8 – 10
Volumen [mm³]	100 – 250
Veränderung	Neu / wachsend (VDT<600 Tage) / stabil (VDT>600 Tage) / abnehmend

SICHERHEITS- UND LEISTUNGSANFORDERUNGEN

Grundlegenden Sicherheits- und Leistungsanforderungen

Die Software ist in der Lage, einmal gelernte strukturelle Veränderungen auf unabhängigen Bilddaten vollautomatisch zu annotieren. Dahinter stehen faltende neuronale Netze (Convolutional Neural Networks – CNNs), die in den Bereich der künstlichen Intelligenz (KI) fallen. Diese Netze werden auf riesigen Datensätzen, die von menschlichen Experten annotiert wurden, trainiert. Die KI wird nur im „eingefrorenen“ Zustand verwendet. Ein Weitertrainieren oder Anpassen der CNN während des Gebrauchs findet nicht statt. Als Ergebnis erstellt die Software einen standardisierten Bericht (Biometrica-Report), der die von der Software erkannten strukturellen Veränderungen visualisiert. Die Erstellung des Berichts orientiert sich an internalen und/oder nationalen Leitlinien oder Standards.

Biometrica erfüllt alle grundlegenden Sicherheits- und Leistungsanforderungen.

Spezifische Sicherheitsanforderungen

Die vollautomatische Annotation von strukturellen Veränderungen (im Vergleich mit dem gesunden Zustand) auf radiologischen Bilddaten durch Biometrica ist nicht schlechter als die von menschlichen Experten (Radiologen).

Spezifische Leistungsanforderungen

Die Durchlaufzeiten (beginnend mit dem Empfang der Bilddaten auf dem zentralen Server und endend mit der Lieferung des Biometrica-Berichts) liegen in 80% aller Fälle unter 30 Minuten.

Klinischer Nutzen

Der klinische Nutzen von Biometrica stellt sich dann ein, wenn neben den grundlegenden Sicherheits- und Leistungsanforderungen auch die Biometrica-spezifischen Sicherheits- und Leistungsanforderungen erfüllt werden. Das bedeutet: Die Biometrica-Auswertung in Form von mind. einen der oben genannten Varianten (Biometrica-Report, Secondary Capture Daten etc.) steht innerhalb von 30 Minuten nach Versand dem Anwender zur Verfügung und die Annotation der strukturellen Veränderungen ist nicht schlechter als jene, die ein Experte erkannt/detektiert hat.

Auf Nachfrage können die Performanz-Daten durch den Hersteller bereitgestellt werden.

Fehlerquellen / Limitationen

Die KI-Algorithmen wurden auf einer großen Menge von qualitätsgesicherten Annotationen (sog. Grundwahrheit) trainiert. Trotz Einhaltung der Akquisitionsparameter können im Routinebetrieb Qualitätsschwankungen in der Bildqualität (Atmungsartefakte, Strahlauflärungsartefakte etc.) auftreten. Diese Qualitätsschwankungen können von den Algorithmen trotz der enormen Bandbreite an Trainingsdaten, die auf Basis von Routedaten erstellt wurden, nicht immer vollständig kompensiert werden und daher zu falschen Ergebnissen (falsch-positive und falsch-negative Ergebnisse) führen. Die Qualität der Auswertungen hängt somit maßgeblich von der Qualität der ihnen zugrunde liegenden Bilddaten ab.

Der gesamte Workflow ab dem Zeitpunkt der abgeschlossenen CT-Aufnahme hängt - zeitlich gesehen - neben der Auswertung von verschiedenen Faktoren wie z.B. der lokalen Netzwerkleistung, der Menge gleichzeitig geschickter Daten, der Größe der Studie/Daten und dem PACS Ihrer Institution sowie dem PACS der jung diagnostics GmbH ab. Daher kann trotz schneller, automatisierter Auswertung, die in der Regel nach Datenerhalt ca. 5-10 Minuten dauert, mit einer Dauer von mind. 30 Minuten bis zum Ergebniserhalt gerechnet werden. Mindestens aber sollte in 80% der Fälle die Auswertung innerhalb von 30 Minuten (nach Dateneingang bei der jung diagnostics GmbH) im Anwender-PACS vorliegen.

Risikobetrachtung

Biometrica stellt in keinerlei Hinsicht ein direktes Sicherheits- oder Gesundheitsrisiko für den Anwender oder den Patienten dar. Indirekte Risiken sind die Identifizierung/Annotation von strukturellen Veränderungen, die nicht pathologisch sind (u.a. Falsch Positive) oder die ausbleibende Identifizierung von Strukturveränderungen, die pathologisch sein könnten.

Das maximal anzunehmende Risiko, das im Rahmen der medizinischen Zweckbestimmung möglich ist, besteht darin, dass die von der Standalone-Software vorgelegte Dokumentation fehlerhaft ist oder nicht den Anforderungen des Nutzers genügt oder aufgrund von Daten- oder Berechnungsfehler nicht erzeugt werden konnte.

Für sonstige technische Anliegen und Fragen zur Sicherheit wenden Sie sich bitte an die jung diagnostics GmbH unter safety@jung-diagnostics.de

INSTALLATION, PACS-INTEGRATION UND WARTUNG

Installation und PACS-Integration

Die jung diagnostics GmbH erhält Studien im DICOM-Format auf zwei unterschiedliche Weisen:

- a. Manueller Versand aus dem PACS oder anderen Modalitäten
- b. Automatischer Versand aus dem PACS oder anderen Modalitäten

Die präferierte Methode wird einmalig bei Installation festgelegt und (im Falle von Punkt b) durch Ihr IT-Personal vor Ort im PACS konfiguriert.

Damit die DICOM-Daten als solche „erkannt“ und durch Biometrica verarbeitet werden können, müssen diese im Standard DICOM-Format (übliches/gängiges Format der im PACS gespeicherten DICOM-Daten) versandt werden.

Biometrica analysiert die Bilddaten nach Vollständigkeit (z.B. Schichtzahl) und nach weiteren notwendigen Voraussetzungen zur Datenauswertung (Kapitel 4.1).

Sollte die Prüfung ergeben, dass die notwendigen Voraussetzungen nicht erfüllt sind, so werden die Bilddaten nicht ausgewertet und es wird kein Ergebnis in Form von DICOM-Daten zur Verfügung gestellt. Ein Fehler-Report / eine Fehlermeldung existiert in diesem Falle nicht.

Wartung

Regelmäßige, kleinere Updates / Software Deployments werden auf Seiten der jung diagnostics GmbH automatisch und ohne Kenntnis des Anwenders durchgeführt, ausgenommen die Änderungen betreffen die Funktionsweisen von Biometrica sowie Software-Versionsänderungen. In diesen Fällen wird der Anwender per E-Mail informiert.