

**Künstliche Neuronale Netze zur automatischen Klassifizierung von CT-Thorax
Bildern in Gesund und Fibrose**

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Medizin
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Mohammad, Qasim
aus Sargodha, Pakistan

Gießen 2020

Aus der

Klinik für Diagnostische und Interventionelle Radiologie und Kinderradiologie,

unter der Leitung von Univ.-Prof. Dr. med. Gabriele A. Krombach,

des Fachbereichs Medizin der Justus-Liebig-Universität Gießen

Gutachter: Prof. Dr. Gabriele A. Krombach

Gutachter: Prof. Dr. Keywan Sohrabi

Tag der Disputation: 07.09.2020

Inhaltsverzeichnis

Einleitung	1
Lungenfibrose.....	1
Definition, Ätiologie und Pathogenese der Lungenfibrose.....	1
Klinische Symptome.....	2
Diagnostik.....	2
Medizinische Bildanalyse.....	3
Bildsegmentierung.....	4
Bildregistrierung.....	4
Visualisierung.....	5
Herausforderungen.....	6
Radiomics.....	7
Historischer Hintergrund.....	7
Definition: Radiomics.....	7
Radiomics: Arbeitsablauf.....	9
Deep Learning.....	12
Lernverfahren.....	13
Hintergrundwissen zum Neuronalen Netz.....	15
Zielsetzung der Arbeit.....	19
Material und Methoden	20
Patienten.....	20
CT-Bildgebung.....	20
TensorFlow.....	21
Arbeitsablauf.....	22
Anzahl der Bilddaten.....	25
PyTorch.....	26
Class-Activation-Maps (CAMs).....	29
Genutzte Computerhardware.....	31

Ergebnisse	32
Diskussion	53
Welchen Vorteil bietet maschinelles Lernen gegenüber regelbasierten Algorithmen?	54
Limitationen dieser Studie.....	57
Strukturierte Befundung	58
Extrahieren der Informationen mittels Natural Language Processing	59
Generelle Aspekte von KI-Methoden	60
Datenqualität.....	63
Klinische Entscheidungsfindung.....	65
Erwartungen an die KI	67
Ausblick und Schlussfolgerung	68
Zusammenfassung	72
Summary	74
Abbildungsverzeichnis	76
Tabellenverzeichnis	78
Literaturverzeichnis	79
Publikationsverzeichnis	89
Ehrenwörtliche Erklärung	90
Danksagung	91
Tabellarischer Lebenslauf	92

Einleitung

Techniken aus dem Bereich der künstlichen Intelligenz, insbesondere des tiefgehenden maschinellen Lernens (Deep Learning), sind Kernbestandteile der jüngsten Entwicklungen auf dem Gebiet der medizinischen Bildgebung. Diese Algorithmen werden für die Segmentierung, Bildrekonstruktion und Klassifikation von Bilddatensätzen verwendet. Da die visuelle Auswertung von Bilddaten oftmals sehr subjektiv und zeitintensiv ist, stellen diese Techniken eine weitere objektive und effiziente Quantifizierungsmethode dar.

Die Möglichkeit einer vollautomatischen Klassifikation von Lungengewebe (in CT-Datensätzen) in „Gesund“ und „Fibrose“ soll im Rahmen dieser Doktorarbeit aufgezeigt werden. Hierzu wurden zum Teil eigenständig Python-Computer-Programme geschrieben und Klassifizierungsmodelle entwickelt.

Lungenfibrose

Definition, Ätiologie und Pathogenese der Lungenfibrose

Die Lungenfibrose ist eine chronische Lungenerkrankung. Hierbei wird gesundes Lungengewebe durch eine veränderte extrazelluläre Matrix ersetzt und die alveoläre Architektur zerstört, was zu einer verminderten Lungencompliance, einem gestörten Gasaustausch und letztlich zu Atemstillstand und Tod führen kann (Richeldi, Collard et al. 2017).

Es zeigt sich eine steigende Inzidenz der idiopathischen Lungenfibrose (IPF). In Europa und in Nordamerika wird von etwa 10 Fällen pro 100 000 Einwohnern pro Jahr berichtet (Hutchinson, Fogarty et al. 2015, Hopkins, Burke et al. 2016). Männer sind häufiger betroffen als Frauen und nur selten tritt die IPF bei Personen unter 50 Jahren auf. Das mittlere Alter bei der Erstdiagnose beträgt in etwa 65 Jahre (Society 2000, Raghu, Weycker et al. 2006, Raghu, Collard et al. 2011). Bei sehr variablem Krankheitsverlauf beträgt die mittlere Überlebenszeit in etwa 2-4 Jahre ab Diagnosestellung (Ley, Collard et al. 2011).

Lange Zeit galt die Lungenfibrose als eine chronische Entzündungsreaktion, die letztendlich zu einem fibrotischen Umbau des Lungengewebes führte. Doch immunsuppressive Therapien mit Prednisolon und Azathioprin erhöhten die Patientensterblichkeit (Selman, King et al. 2001, Network 2012). Nach aktuellen Erkenntnissen wird die Lungenfibrose als eine allgemeine Folge multifaktorieller und

wiederholter Mikroverletzungen des Alveolarepithels angesehen. Diese wiederholten Verletzungen tragen zu einer gestörten Epithel-Fibroblasten-Kommunikation bei, welches zu der Induktion von matrixproduzierenden Myofibroblasten führt (Richeldi, Collard et al. 2017). Umweltfaktoren, wie zum Beispiel die Inhalation von Partikeln tragen maßgeblich zur Pathogenese der idiopathischen Lungenfibrose bei. Insbesondere spielt der Nikotinabusus eine wesentliche Rolle bei der Entwicklung einer Lungenfibrose (Baumgartner, Samet et al. 1997).

Darüber hinaus scheint die genetische Prädisposition eine pathogenetische Rolle einzunehmen. Mehrere Gene wurden identifiziert, die zu einem gesteigerten Risiko führen (Nogee, Dunbar et al. 2001, Armanios, Chen et al. 2007, Tsakiri, Cronkhite et al. 2007, Wang, Kuan et al. 2009, Noth, Zhang et al. 2013, Stuart, Choi et al. 2015).

Klinische Symptome

Dyspnoe und Husten sind wichtige, jedoch unspezifische Symptome bei Patienten mit einer Lungenfibrose. Deshalb werden oft diese Erstsymptome bei Erstvorstellung beim Hausarzt im Rahmen des natürlichen Alterungsprozesses aufgefasst oder als eine Herz-Kreislaufstörung gewertet. In der körperlichen Untersuchung fällt ein bilaterales Inspirationsknistern (King, Tooze et al. 2001) und im Lungenfunktionstest fallen eine reduzierte Totalkapazität und ein reduzierter Gasaustausch auf (Raghu, Weycker et al. 2006).

Diagnostik

Neben der konventionellen Röntgenübersichtsaufnahme des Thorax spielt die Computertomographie eine Schlüsselrolle bei der Detektion kleinster pulmonaler Veränderungen. Die hochauflösende Computertomographie des Thorax (HRCT) ermöglicht eine detaillierte Beurteilung des Lungenparenchyms ohne Überlagerungsphänomene. Schichten von etwa 1 mm dienen zur besseren Beurteilung des genauen Ausmaßes und der Verteilung der Lungenfibrose. Typische Muster, die auf eine Lungenfibrose hindeuten, sind Traktionsbronchiektasien und subpleurale, bilaterale Honigwabenmuster mit basaler Verteilung. Die Honigwaben sind als Cluster zystischer Lufträume mit Durchmessern von etwa 3 -10 mm zu erkennen. Mosaikmuster, Milchglastrübungen und noduläre Verdichtungen legen eine andere Diagnose nahe (Hunninghake, Lynch et al. 2003, Hansell, Bankier et al. 2008, Travis, Costabel et al.

2013). Sollten in der HRCT keine eindeutigen Merkmale zu identifizieren sein, wird eine operative Lungenbiopsie in Erwägung gezogen. Bei Patienten unter 65 Jahren und mit wenigen Komorbiditäten beträgt bei elektiven Eingriffen die Sterblichkeit etwa 2 % in den folgenden 30 Tagen (Hutchinson, Fogarty et al. 2016). Bei älteren Patienten mit zahlreichen Komorbiditäten (Charlson Score >1) (Richeldi, Collard et al. 2017) und klinisch signifikanten physiologischen Beeinträchtigungen sollte eine Lungenbiopsie kritisch hinterfragt werden (Hutchinson, Fogarty et al. 2016, Richeldi, Collard et al. 2017). Klinische Symptome, die radiologische Bildgebung und die Histopathologie spielen bei der Diagnose der Lungenfibrose eine entscheidende Rolle. Ein multidisziplinärer Dialog zwischen Ärzten erhöht die diagnostische Treffsicherheit (Flaherty, King et al. 2004). Die Genauigkeit der Diagnose hängt somit stark von der Expertise der behandelnden Ärzte ab. Es zeigt sich eine hohe Inter- und Intraobservervariabilität (Flaherty, Andrei et al. 2007). Der Wunsch nach einer effizienten und objektiven Klassifizierungsmethode ist groß.

Medizinische Bildanalyse

Aufgrund der anhaltenden technologischen Fortschritte auf dem Gebiet der medizinischen Bildgebung steigt die Menge der zu befundenen Daten. Die zeitnahe und qualitative Analyse dieser Bilddaten ist zu einer enormen Herausforderung geworden. Gleichzeitig stellt die Einführung der digitalen Bildarchivierungs- und Kommunikationssysteme (PACS) und deren Integration innerhalb des gesamten Krankenhausinformationssystems (KIS) sicher, dass große medizinische Bild-Datenbanken und zugehörige relevante medizinische Informationen von Patienten (einschließlich demographischer Daten, klinischer Befunde, Labordaten und Pathologie) einheitlich aufgebaut werden. Sofern technische Herausforderungen und ethische Fragen des Datenschutzes angemessen beantwortet werden, können solche Datenbanken mehr und mehr zu Forschungszwecken eingesetzt werden.

Die medizinische Bildanalyse beschäftigt sich unter anderem mit Messungen in medizinischen Bildern. Manuelle Messungen in großen medizinischen 3D-Bilddatensätzen sind nicht nur mühsam und zeitaufwendig, sondern unterliegen auch einer erheblichen Intra- und Interobserver-Variabilität, die die Objektivierbarkeit der daraus abgeleiteten Befunde einschränkt. Es besteht daher ein großer Bedarf an effizienteren, zuverlässigen und gut validierten automatisierten oder halbautomatischen Methoden für die medizinische Bildanalyse. Dies ist der erste Schritt zur computergestützten Bildinterpretation in der klinischen Routine. Welche Informationen

aus den Bildern quantifiziert werden müssen, ist sehr anwendungsspezifisch und erheblich von der Fragestellung abhängig. In der medizinischen Bildanalyse geht es häufig um die Quantifizierung bestimmter geometrischer Merkmale, die Beurteilung anatomischer Veränderungen im Laufe der Zeit (z.B. im Rahmen von Verlaufsuntersuchungen) oder die Erkennung und Charakterisierung morphologischer Variationen (z.B. Normvariante versus Pathologie). Die Analyse der 3D-Form- und insbesondere der Formvariabilität stellt eine besondere Herausforderung in der medizinischen Bildanalyse dar. Neben der Morphometrie ist die Quantifizierung von lokalen oder regionalen Kontrasten oder Kontrastunterschieden in vielen Anwendungen von Bedeutung, insbesondere in der funktionellen Bildgebung, wie z.B. fMRI, PET oder MR-Diffusions- und Perfusionsbildgebung.

Bildsegmentierung

Bei der Bildsegmentierung wird zwischen den zu einem bestimmten Objekt gehörenden und den nicht zu diesem Objekt gehörenden Bild-Voxeln unterschieden. Die Bildsegmentierung ist eine Voraussetzung für die Quantifizierung der geometrischen Eigenschaften eines Objekts, insbesondere seines Volumens oder seiner Form. Die Bildsegmentierung kann auf unterschiedliche Weise durchgeführt werden: grenzwertbezogen (threshold) durch die Beschreibung der Kontur oder Oberfläche des Objekts in einem (2D) oder mehreren (3D) Bildebenen; regionenbezogen durch die Gruppierung von Voxeln; oder voxelbezogen durch die Zuordnung jedes Voxels im Bild zu einem bestimmten Objekt. Die genaue 3D-Segmentierung von komplex geformten Objekten in medizinischen Bildern wird in der Regel durch die begrenzte Auflösung der Bilder und durch die Tatsache, dass die Auflösung oft nicht isotrop ist, erschwert. Um fehlende Informationen in den Daten zu ergänzen, ist in der Regel eine Interpolation erforderlich. Im klinischen Alltag sind präzise 3D-Messungen (z.B. Volumetrie von Lungenläsionen) sehr mühsam und zeitaufwendig, so dass oft vereinfachte, ungefähre 2D- oder 1D-Messungen stattfinden.

Bildregistrierung

Die Bildregistrierung beschäftigt sich mit der räumlichen Beziehung zwischen verschiedenen Bildern. Diese räumliche Beziehung wird anhand des Bildinhalts selbst durchgeführt (Bild-Matching) (Hajnal and Hill 2001). Unterschiedliche Bilder, die zu

unterschiedlichen Zeitpunkten (z.B. vor und nach der Behandlung) oder mit unterschiedlichen Modalitäten aufgenommen wurden (z.B. CT-, MRT-, PET-Untersuchungen) enthalten oft komplementäre Informationen, die zusammengeführt und gemeinsam analysiert werden müssen. Zudem wird die Bildregistrierung benötigt, um unterschiedliche Patientenpositionierungen im Scanner und um Gewebedeformationen zwischen verschiedenen Zeitpunkten auszugleichen. Nach der korrekten Registrierung können die Bilder durch präzise Überlagerung gemeinsam und parallel durchgearbeitet werden.

Visualisierung

Die aus den Bildern extrahierten Informationen sollten idealerweise zur Unterstützung der Diagnose und Therapieplanung optimal dargestellt werden, d.h. so, dass die richtige Interpretation aller relevanten Bilddaten für den Anwender maximal erleichtert wird. Für medizinische 3D-Bilder ist die multiplanare 2D-Visualisierung nicht gut geeignet, um strukturelle Beziehungen innerhalb und zwischen Objekten zu beurteilen- hierfür sind echte 3D-Visualisierungsansätze zu bevorzugen, wie zum Beispiel Oberflächen- oder Volumen-Rendering.

In klinischen Anwendungen, wie beispielsweise der bildbasierten Operationsplanung oder der bildgesteuerten intraoperativen Navigation müssen Werkzeuge bereitgestellt werden, um Objekte in der 3D-Darstellung nach Bedarf zu bearbeiten (z.B. Messungen durchzuführen) oder um virtuelle Objekte einzufügen (z.B. Implantate).

Bildsegmentierung, -registrierung und -visualisierung sind nicht absolut getrennte Säulen der medizinischen Bildanalyse. Im Gegenteil, sie sind in der Regel miteinander verknüpft. Eine optimale Lösung für eine bestimmte Fragestellung kann nur durch eine gemeinsame Betrachtung von Segmentierung, Registrierung und Visualisierung erreicht werden. Beispielsweise kann die Bildregistrierung für die Bildsegmentierung hilfreich werden, indem das zu segmentierende Bild mit einem ähnlichen Bild (z.B. von einem anderen Patienten oder einer Atlasvorlage), das zuvor segmentiert wurde, verglichen wird. Umgekehrt kann die Bildregistrierung davon profitieren, dass bereits eine vorherige Segmentierung ähnlicher Strukturen stattgefunden hat. Die gemeinsame Visualisierung unterschiedlicher Bilder, die z. B. prä- und intraoperativ aufgenommen wurden, setzt voraus, dass die Registrierungsprobleme zwischen diesen Bildern, die in unterschiedlichen Koordinatensystemen liegen, gelöst sind. Darüber hinaus muss bei bildgestützter Behandlungen der präoperative Behandlungsplan während des Eingriffs auf den

Patienten übertragen und die intraoperative Position der Instrumente in den Bildern verfolgt werden. Dieses Registrierungsproblem erfordert in der Regel die Installation zusätzlicher Hardware im Behandlungsraum (z.B. ein optisches Trackingsystem).

Herausforderungen

Medizinische Bilder sind zum größten Teil dreidimensional. Die 3D-Natur der Bilder liefert zusätzliche Informationen, aber auch eine zusätzliche Dimension der Komplexität. Anstatt die Daten scheibenweise in 2D zu verarbeiten, ist die 3D-Verarbeitung in der Regel effektiver, da sie die räumlichen Zusammenhänge berücksichtigt. Medizinische Bilder basieren auf unterschiedlichen physikalischen Prinzipien und die Beurteilung bzw. Interpretation dieser Bilder wird durch die Mehrdeutigkeit erschwert. Diese Mehrdeutigkeit wird hauptsächlich durch die Grenzen des Bildaufnahmeprozesses, insbesondere durch die begrenzte Auflösung, dem beschränkten Kontrastspektrum, das Rauschen bzw. das Vorhandensein von Artefakten bestimmt. Viele Fragestellungen erfordern die Analyse von komplementären Informationen, die von mehreren Bildern geliefert werden, um beispielsweise anatomische und funktionelle Informationen zu korrelieren, um Veränderungen im Laufe der Zeit oder Unterschiede zwischen den Patienten zu beurteilen. Objekte, die in medizinischen Bildern von Interesse sind, sind typischerweise anatomische Strukturen, Fremdkörper, z.B. Implantate, Normvarianten oder pathologische Läsionen. Anatomische Strukturen können komplexe Formen aufweisen, wie z.B. die Gehirn- und Herzkranzgefäße. Solche komplexen Formen lassen sich nicht immer einfach durch mathematische Modelle beschreiben. Zudem weisen anatomische Strukturen eine große Formvariabilität auf. Eine absolute Abbildungsgenauigkeit ist aufgrund fehlender Grundwahrheiten oft nicht möglich. Als Alternative könnte ein Modell (Phantom), welches die relevanten Objekte abbildet, genutzt werden. Ein generalisiertes Hardware-Phantom stellt sich jedoch in der Regel als ein ziemlich starres Modell dar und lässt sich nur schwierig an die anatomische Vielfalt anpassen. Es fehlt die Flexibilität. Ein Software-Modell in Kombination mit einem Simulator bietet hingegen mehr Flexibilität und Anpassungsfähigkeit. Ehe solche Modelle jedoch im klinischen Alltag eingesetzt werden können, müssen diese auf Genauigkeit, Präzision, Konsistenz und Robustheit überprüft werden.

Radiomics

Historischer Hintergrund

Der Gedanke bildgebende Merkmale zur Klassifikation von Bildern einzusetzen, hat seine Wurzeln in den 1960er Jahren (Doi and graphics 2007). Der Einsatz dieser Merkmale fokussierte sich hauptsächlich auf die computergestützte Erkennung oder Diagnose (CAD) (Tang, Rangayyan et al. 2009, Giger, Karssemeijer et al. 2013). CAD verwendet eine Reihe von quantitativen Bildmerkmalen, die die geometrische Struktur, die Intensitätsverteilung und die Textur einer „Region of Interest“ (ROI) beschreiben. Diese werden anschließend von statistischen oder maschinellen Klassifikationsverfahren verwendet, um beispielsweise zwischen benignen und malignen Läsionen zu unterscheiden. Das Ergebnis kann anschließend von Radiologen als Zweitmeinung bei der Erkennung von Läsionen und bei diagnostischen Entscheidungen verwendet werden (Doi and graphics 2007, Tang, Rangayyan et al. 2009, Giger, Karssemeijer et al. 2013, Dhungel, Carneiro et al. 2017, Summers 2017). CAD wird hauptsächlich in zwei Hauptgruppen unterteilt, die computergestützte Erkennung (CADe) und die computergestützte Diagnose (CADx). CADe beschäftigt sich mit der Erkennung, Lokalisierung und Segmentierung von Läsionen in medizinischen Bildern. CADx beschäftigt sich dagegen mit der Diagnose, z.B. mit der Unterscheidung zwischen gutartigen und bösartigen Läsionen (Giger, Karssemeijer et al. 2013, El Naqa, Li et al. 2015). Somit bestehen CAD-Algorithmen im Allgemeinen aus zwei Stufen, der Erkennung und Klassifizierung suspekter Läsionen. In der ersten Stufe werden Texturmerkmale aus ROIs extrahiert. Da viele Merkmale extrahiert werden können, enthalten CAD-Systeme häufig Algorithmen zur Merkmalsauswahl, um die Merkmale auszuwählen, die am signifikantesten zur diagnostischen Genauigkeit beitragen (Summers 2017). In der zweiten Stufe wird ein binärer Klassifikator angelehrt (Elter and Horsch 2009).

Definition: Radiomics

Der Begriff Radiomics stammt von „Radio“ ab, was sich auf die Gewinnung medizinischer Bilder durch den Einsatz von Strahlung (z.B. Röntgen, CT, MRI) bezieht. Das Suffix „omics“ wurde erstmals im Begriff Genomics verwendet, um die Kartierung

des menschlichen Genoms zu bezeichnen. Radiomics hat ihre Wurzeln in der CAD und kombiniert die quantitative Analyse von radiologischen Bildern mit Methoden des maschinellen Lernens (Parekh, Jacobs et al. 2016). Gegenüber CAD sind jedoch zwei Aspekte in der Radiomics neu: die Anzahl der beteiligten Bildmerkmale ist in der Radiomics deutlich höher und die Merkmale, die aus radiologischen Bildern extrahiert werden, stehen im engen Zusammenhang mit biologischen oder klinischen Endpunkten (Gillies, Kinahan et al. 2016). Radiomics ist eine allgemeine Wissenschaft, die auf viele biomedizinische Bereiche angewendet werden kann. Es ist derzeit anerkannt, dass solide Tumore nicht aus einer homogenen Einheit bestehen, sondern aus mehreren klonalen Subpopulationen von Krebszellen, die eine beträchtliche räumliche und zeitliche Variabilität aufweisen, die möglicherweise wertvolle Informationen über die Aggressivität von Tumoren liefern könnte (Vallières, Freeman et al. 2015). Quantitative Bildmerkmale, auch „radiomische Merkmale“ genannt, könnten Informationen über Intensität, Form, Größe oder Volumen und Textur des Tumorphänotyps liefern, die sich mit denen aus klinischen Berichten ergänzen. Auf Biopsieergebnissen basierte Auswertungen sind nicht nur invasiv, sondern ermöglichen auch eine begrenzte Tumorcharakterisierung, da die extrahierte Probe nicht immer die gesamte Population von Tumorzellen repräsentiert. Radiomics umgeht dies, indem die umfassende dreidimensionale Tumormasse mit Hilfe von bildgebenden Informationen beurteilt wird (Aerts, Velazquez et al. 2014). Verschiedene bildgebende Verfahren (z.B. MRT, CT, PET, Ultraschall) werden als Grundlage für die Extraktion dieser Merkmale verwendet (Lambin, Rios-Velazquez et al. 2012). Eine Sammlung von Merkmalen, die einen prognostischen und/oder prädiktiven Wert haben, wird oft als „radiomische Signatur“ bezeichnet. Die grundlegende Hypothese ist, dass die quantitative Analyse von Tumoren durch eine große Anzahl von radiomischen Merkmalen wertvolle diagnostische, prognostische oder prädiktive Informationen liefern kann (Kumar, Gu et al. 2012, Panth, Leijenaar et al. 2015). Das Ziel ist es, diese Informationsquellen zu erforschen und zu nutzen, um ein diagnostisches, prädiktives oder prognostisches radiomisches Modell (Signaturen) zu entwickeln, um personalisierte klinische Entscheidungen zu unterstützen und die individualisierte Behandlungsauswahl zu verbessern (Gillies, Kinahan et al. 2016).

Radiomics: Arbeitsablauf

Der Prozess bis zur Erstellung einer radiomischen Signatur mit prognostischem/prädiktivem Wert hat vier Stufen (Kumar, Gu et al. 2012, Coroller, Grossmann et al. 2015, Parmar, Grossmann et al. 2015, Parmar, Leijenaar et al. 2015, Gillies, Kinahan et al. 2016): Die erste Stufe umfasst die Bildaufnahme, gefolgt von einer automatischen oder manuellen Segmentierung der ROI. Anschließend werden die Bildmerkmale extrahiert und ein mathematisches Modell wird erstellt.

Bildaufnahme

Die am weitesten verbreitete bildgebende Methode, die bisher mittels Radiomics ausgewertet worden ist, ist die CT. Mit dieser Untersuchung können die Dichte, Form und Textur von Tumoren und Lymphknoten beurteilt werden (Aerts, Velazquez et al. 2014, Giesel, Schneider et al. 2016, Huang, Liu et al. 2016). Die CT hat auch das Potenzial, die Größe und die Größenveränderung während und nach einer Therapie zu bestimmen, um das Ansprechen auf eine Behandlung rechtzeitig einschätzen zu können. Texturmerkmale, die aus der CT extrahiert werden können, stehen in einigen Fällen in Zusammenhang mit dem Tumorstoffwechsel (Ganeshan, Abaleke et al. 2010) und der Histopathologie (Ganeshan, Goh et al. 2013). Bei einigen Tumoren stehen zudem radiologische CT-Merkmale wie Nekrose und Verkalkung in Zusammenhang mit Genmutationen, die eher mit einem fortgeschrittenen Stadium assoziiert sind (Karlo, Di Paolo et al. 2014). Die PET wird zum Nachweis und zum Staging von Tumoren eingesetzt. Sie ist eine wichtige Untersuchung in Radiomics und liefert Informationen über die funktionelle Natur des Tumors und eventuell seiner zugrunde liegenden Tumorbiologie (El Naqa and imaging 2014). Seit Ende der 2000er Jahre ist die Texturanalyse in der PET weit verbreitet (Hatt, Tixier et al. 2017). In einer multizentrischen Studie wurde festgestellt, dass die PET-Texturmerkmale zur Vorhersage des Gesamtüberlebens bei Lungentumoren eingesetzt werden konnten.

Die MRT liefert kontrastreiche strukturelle und funktionelle Informationen zur Charakterisierung von Weichgewebe. Eine der ersten Radiomics-Studie, die auf der MRT basierte, zeigte, dass Bereiche mit erhöhter Neovaskularität und Zellproliferation mit den Biopsieergebnissen korreliert werden konnten (Earnest 4th, Kelly et al. 1988). Radiomische Merkmale konnten erfolgreich genutzt werden, um Prostatakrebs von gesundem Gewebe zu differenzieren (Cameron, Khalvati et al. 2015, Khalvati, Wong et al. 2015). Die dynamische kontrastverstärkte MRT stellt die Konzentration eines

injizierten Gadolinium-Kontrastmittels über die Zeit dar (Khalvati, Wong et al. 2015). Unter Ausnutzung der Anreicherungskinetik kann der zeitliche Verlauf der Signalintensität innerhalb der Läsion bei der Interpretation von Läsionen zur Bestimmung der Wahrscheinlichkeit einer Malignität herangezogen werden (Bhooshan, Giger et al. 2010). Radiomische Merkmale konnten zwischen den Expressionsstatus des Östrogenrezeptors (ER) und des Progesteronrezeptors (PR) bei Mammakarzinom unterscheiden (Li, Zhu et al. 2016).

Segmentierung der „Region of Interest (ROI)“

Die Extraktion von radiomischen Merkmalen wird normalerweise auf eine spezifische ROI angewendet. Diese muss häufig zuerst manuell von einem erfahrenen Radiologen oder Radioonkologen abgegrenzt werden (Aerts, Velazquez et al. 2014, Vallières, Freeman et al. 2015, Wu, Parmar et al. 2016). Es wurden mittlerweile eine Reihe von Algorithmen zur automatischen und halbautomatischen Segmentierung entwickelt. „Region Growing“ ist eine halbautomatische Methode, die häufig bei der Segmentierung angewendet wird. Hierbei werden homogene Bildelemente zu Regionen gruppiert (Elter and Horsch 2009). Beginnend mit einer beliebigen Wahl der Anfangsregion, wird diese mit den Nachbarbildpixeln anhand eines Kriteriums verglichen. Trifft das Kriterium zu, werden die Nachbarpixeln als zu dieser Region zugehörig definiert. Dieser Ansatz wird iterativ wiederholt, bis keine Nachbarpixel mehr hinzugenommen werden können. Das Bild kann somit in unterschiedliche Regionen eingeteilt werden. Die automatische Segmentierung von Brusttumoren in der MRT wird seit mehr als 10 Jahren eingesetzt, wobei die Ergebnisse mit der manuellen Segmentierung durch einen Radiologen vergleichbar sind (Chen, Giger et al. 2006, Li, Zhu et al. 2016). In einer Brustkrebs-Radiomics-Studie (Wang, Kato et al. 2015) wurden die betroffene Brust, der Tumor und das umgebende Parenchym halbautomatisch auf MRT-Datensätzen segmentiert (Tixier, Hatt et al. 2014). Im Jahr 2014 wurde ein vollautomatisches Segmentierungswerkzeug für Hirntumore (BraTumIA) und deren Kompartimente vorgeschlagen (Porz, Bauer et al. 2014). Die automatische Prostatakrebs-Segmentierung wurde mit Hilfe der multiparametrischen MRT durchgeführt. Die Segmentierung beruhte vornehmlich auf T2-gewichtete Sequenzen und auf DWI Protokollen (Cameron, Khalvati et al. 2015, Khalvati, Wong et al. 2015).

Extraktion der Radiomics-Merkmale

In der Radiomics werden große Mengen quantitativer Bildmerkmale extrahiert und analysiert. Mit dem Begriff „Signatur“ ist eine Sammlung von Merkmalen gemeint, die prognostischen und/oder prädiktiven Wert haben. Radiomische Merkmale werden nach der Bildvorverarbeitung und Filterung extrahiert (Coroller, Grossmann et al. 2015). Diese Merkmale können in diverse Kategorien unterteilt werden. Einige kürzlich eingeführte Kategorien beinhalten fraktale und Fusionsmerkmale, die nur für multimodale Datensätze anwendbar sind. Größen- und Formmerkmale, wie z.B. Volumen, Oberfläche, 2D- und 3D-Maximaldurchmesser beschreiben die Größe der ROI. Das morphologische Erscheinungsbild des Randes wird durch die Randschärfe, die das Mittel des Bildgradienten am Läsionsrand ist, beschrieben (Li, Zhu et al. 2016). Ein Histogramm beschreibt Merkmale, die mit der Verteilung der Intensitäten der Voxel innerhalb der ROI zusammenhängen. Ebenfalls kann der zeitliche Verlauf einer Signalintensität innerhalb der Läsion nach Kontrastmittelinjektion betrachtet werden. Der Begriff Bildtextur bezieht sich auf die wahrgenommene oder gemessene räumliche Variation der Intensitätsstufen, die als Graustufenskala visualisiert werden kann (Gnep, Fargeas et al. 2017). Fusionsmerkmale sind mit multimodalen Bilddatensätzen verbunden, die durch Registrierung geometrisch ausgerichtet werden können.

Machine Learning: Erstellen von Radiomics-Klassifikatoren

Das Ziel des Einsatzes von Radiomics ist es, ein prognostisches und/oder prädiktives mathematisches Modell zu entwickeln. Hierbei werden die radiomischen Merkmale analysiert und daraus wird eine Funktion abgeleitet. Diese kann anschließend bestenfalls das Patientenoutcome oder den Tumorphänotyp vorhersagen (Parmar, Grossmann et al. 2015). Aus jedem Patientenbild kann eine große Anzahl von Merkmalen extrahiert werden, typischerweise im Bereich von einigen hundert- bis hin zu tausenden (Aerts, Velazquez et al. 2014). Nicht alle Merkmale wären jedoch für einen Klassifikator sinnvoll. Die Auswahl der Merkmale ist erheblich von der Fragestellung abhängig. Bisher wurden viele Methoden verwendet, um Klassifizierungsmodelle aus radiomischen Merkmalen zu erstellen (Tang, Rangayyan et al. 2009). Zu den beliebtesten überwachten Klassifikatoren gehört, möglicherweise aufgrund ihrer Einfachheit, die logistische Regression (Vallières, Freeman et al. 2015, Yoon, Sohn et al. 2015). Künstliche Neuronale Netze (KNNs), eine der klassischen Methoden des maschinellen Lernens (Elter and Horsch 2009), werden auch bei Einsatz von Radiomics zur überwachten Klassifikation eingesetzt (Wu, Parmar et al. 2016). Bei unüberwachten

Klassifikatoren liegen die Anlern Daten in zuvor unbekannte bzw. undefinierte Gruppen vor (Parekh, Jacobs et al. 2016). Die Patienten werden auf der Basis einer Abstandsmetrik gruppiert.

Deep Learning

In den letzten Jahren haben Deep Learning Methoden einen großen Einfluss auf verschiedenste Lebensbereiche genommen. Sie haben beispielsweise zu erheblichen Verbesserungen auf den Gebieten der Sprach- und Bilderkennung geführt (Dahl, Yu et al. 2011, Krizhevsky, Sutskever et al. 2012). Auch auf den Gebieten, auf denen der menschliche Verstand bisher geherrscht hat, zeigen Deep Learning Algorithmen eine enorme Effizienz und zeichnen sich durch ihre hohe Präzision aus. Es gibt wahrscheinlich drei Gründe für den aktuellen Erfolg von Deep-Learning-Algorithmen: die Verfügbarkeit von großen Datenmengen, die erhöhte und effizientere Rechenleistung und die schnelle, qualitative Entwicklung von Algorithmen. Diese drei Hauptgründe sind stark miteinander verknüpft: Die Verfügbarkeit großer Bilddatensätze und die Rechenleistung ermöglichten es, die Stärke der Deep-Learning-Algorithmen zu demonstrieren. Diese Erfolge regten wiederum die Verwendung noch größeren Datensätze und die Entwicklung neuer Algorithmen an. Moderne Grafikkarten (GPUs), die zur schnellen Datenverarbeitung eingesetzt werden, haben die Rechenzeiten drastisch reduziert und neue große wissenschaftliche Erkenntnisse mit sich gebracht. So sind diese Algorithmen in der Lage menschliche Spieler in Go und ATARI zu schlagen und selbständig neue Bilder und Musik zu generieren (Mnih, Kavukcuoglu et al. 2015, Silver, Huang et al. 2016, Briot, Hadjeres et al. 2017, Tan, Chan et al. 2017). In der medizinischen Bildgebung zeichnen sich diese Algorithmen durch eine hohe Zuverlässigkeit aus. Forscher aus den USA entwickelten Modelle, die anhand von histologischen Bildern Adenokarzinome der Lunge auf dem Niveau erfahrener Pathologen in Subtypen einteilen können (Wei, Tafe et al. 2019). In der Dermatologie ist eine neuronale Netz-Software auf der Grundlage von einfachem Fotomaterial in der Lage, geringfügig besser zwischen benignen und malignen Hauttumoren zu differenzieren, als es erfahrene Dermatologen können (Esteva, Kuprel et al. 2017). In der Radiologie sind 3D-Neuronale Netze in der Lage Lungenrundherde zuverlässig zu detektieren (Jung et

al. 2018). Deep Learning ist eine Subklasse des maschinellen Lernens. Das maschinelle Lernen ist selbst eine Unterklasse der künstlichen Intelligenz (Abbildung 1).

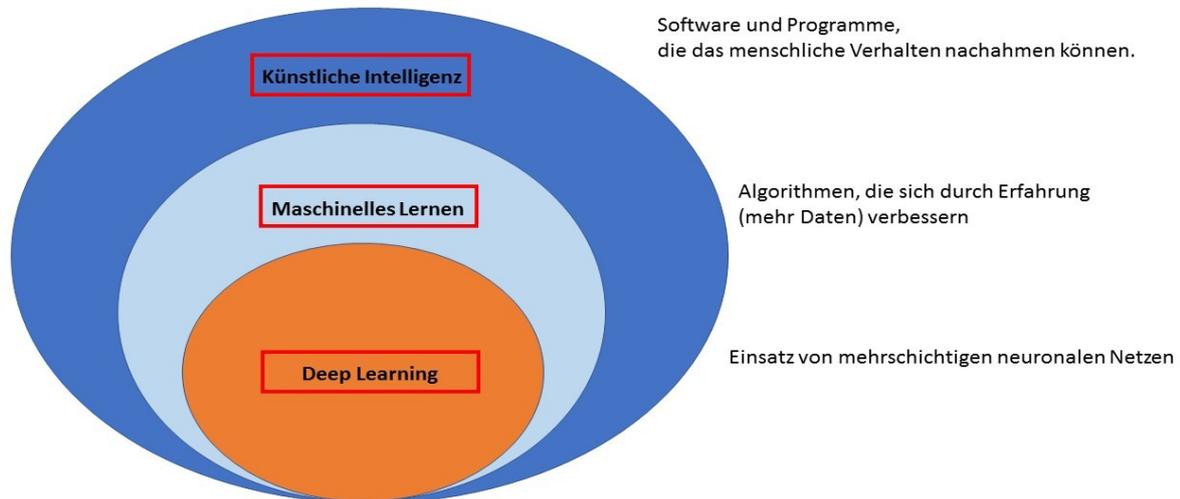


Abbildung 1: Deep Learning als Subklasse

Lernverfahren

Hauptsächlich wird zwischen überwachtem und unüberwachtem Lernen unterschieden. Das überwachte Lernen findet zum Großteil in der Klassifizierung von Daten Anwendung. Zunächst werden die Anlern Daten entsprechend ihrer Zuordnungsgruppe beschriftet. Es folgt die Anlernphase, in der nach Mustern in der jeweiligen Gruppe gesucht wird. Am Ende des Anlernens ist eine Verallgemeinerung möglich und es können somit neue Daten klassifiziert werden (Abbildung 2) (Deo Rahul 2015, Fabris, Magalhães et al. 2017).

Das unüberwachte Lernen wird vorwiegend eingesetzt, um große Datenmengen in Gruppen zu sortieren. Die Daten liegen hierbei unbeschriftet vor (Le 2013, Deo Rahul 2015). Der Algorithmus erkennt Muster und teilt die Daten in Gruppen ein (Abbildung 3). Dies wird als „Clustering“ bezeichnet.

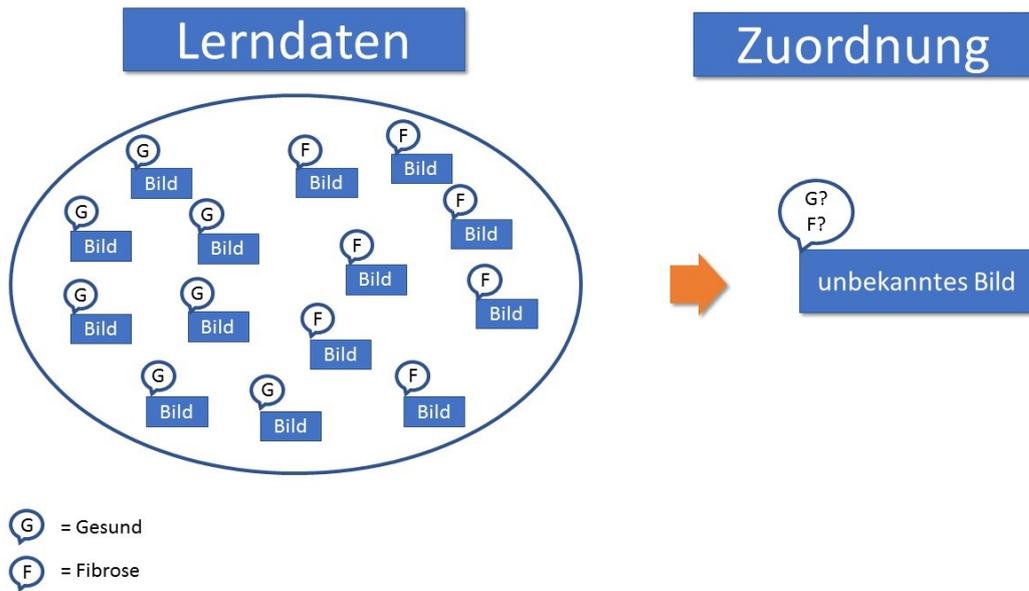


Abbildung 2: Supervised Learning als Lernverfahren

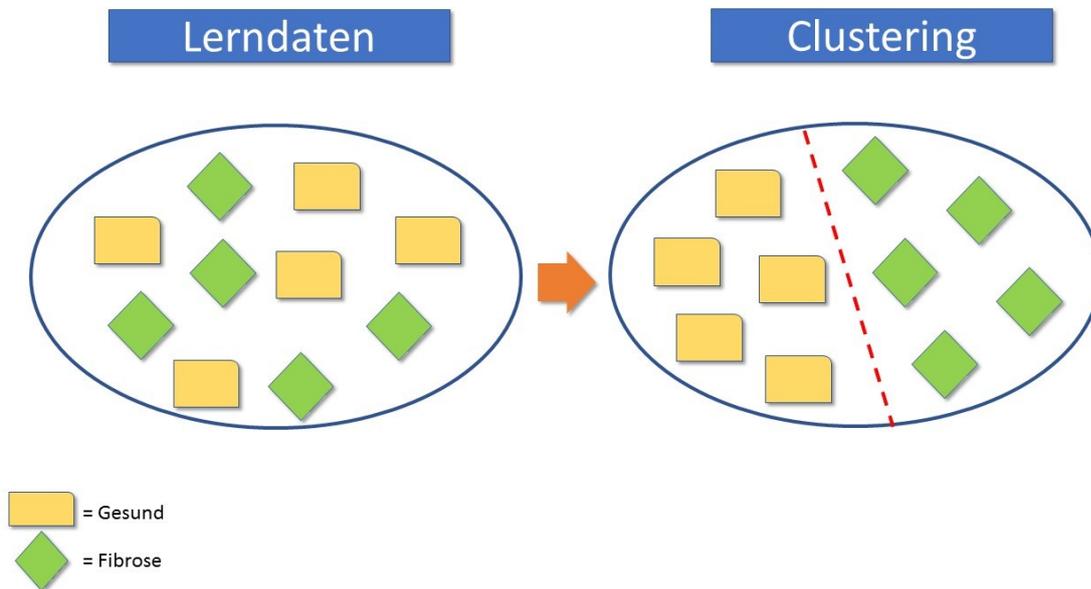


Abbildung 3: Clustering – automatische Gruppenzuordnung

Hintergrundwissen zum Neuronalen Netz

Aufgaben, wie Gesichts-, Sprach-, und Bilderkennung fallen uns Menschen nicht schwer. Wir können sehr schnell Gesichter erkennen und zuordnen, auch wenn der Raum nicht sehr gut beleuchtet ist. Auch wenn Menschen einen neuen Haarschnitt haben oder andere Kleider tragen, bleibt für uns die Gesichtserkennung eine leichte Aufgabe. Die Spracherkennung ist für uns ebenfalls keine Herausforderung. So können wir andere Menschen verstehen, die wir zuvor nicht gehört haben. Ihre Stimme ist für uns fremd, dennoch fällt uns das Verstehen (angenommen bei gleicher Sprache und gleichem Dialekt) nicht schwer. Der Mensch ist in der Lage sich an neue Situationen sehr schnell anzupassen und hat das Potential zu lernen. Die Leistungsfähigkeit des menschlichen Gehirns ist enorm und sehr faszinierend. In der Informatik beseht deshalb das Bemühen, die Funktionsweise des menschlichen Gehirns nachzuahmen (Jain, Jianchang et al. 1996).

Neurone sind die kleinste funktionelle Einheit des Nervensystems und leiten elektrische Signale sowohl aus der Körperperipherie an das Gehirn, als auch umgekehrt. Vereinfacht gesehen bestehen biologische Neurone aus einem Zellkörper (beinhaltet den Nucleus und weitere Zellbestandteile), Dendriten und Axonen. Dendriten sind dünne Fortsätze des Zellkörpers, fungieren als Empfänger und nehmen somit Reize aus der Umgebung auf. Diese aufgenommenen Reize gelangen über Axone an die nächste Zelle (Abbildung 4).

Neuronen sind mit Millionen von anderen Neuronen verknüpft und stellen somit ein Netz dar. Dieses Netz wird als Neuronales Netz bezeichnet.

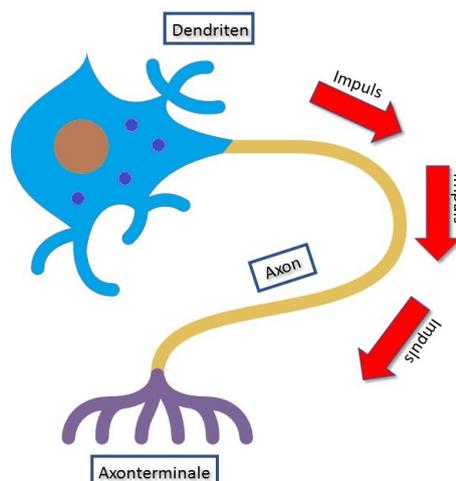


Abbildung 4: vereinfachte Darstellung eines Neurons

Ein komplexes Zusammenspiel aus dem Gehirn und den Sinneseindrücken lässt das Sehen und das Erkennen von Objekten einfach erscheinen: Für Menschen bedarf es keiner oder nur sehr wenig Anstrengung eine Katze von einem Hund zu unterscheiden oder das Gesicht eines Menschen zu erkennen. Dies erscheint aber nur deshalb einfach, weil das Gehirn exzellent darin ist, Bilder zu verstehen. Daher ist es eine umso größere Herausforderung diese Aufgabe mit einem Computer zu realisieren.

In den letzten Jahren hat es im Bereich des maschinellen Lernens enorme Fortschritte gegeben. Insbesondere haben künstliche Neuronale Netze, vor allem das tiefe neuronale Faltungsnetz (Convolutional Neural Network (CNN)), eine hervorragende Leistung bei schwierigen visuellen Erkennungsaufgaben gezeigt (Krizhevsky, Sutskever et al. 2012, Farabet, Couprie et al. 2013, Tompson, Jain et al. 2014, LeCun, Bengio et al. 2015).

CNNs sind von biologischen Prozessen inspiriert und entwickelt worden, um Muster direkt aus Bildern zu erkennen. Zuerst findet eine Merkmalsextraktion statt, woraufhin die Klassifizierung folgt. Faltungsschichten (Convolutional Layers), Aktivierungsschichten (Activation Layers), Poolingsschichten (Pooling Layers) und vollständig verbundene Schichten (Fully Connected Layers) gehören zu einem klassischen CNN. Jedes Neuron der Faltungsschicht ist mit einem kleinen Eingangsbereich des Bildes verbunden. Dieser Eingangsbereich entspricht dem Empfangsfeld des menschlichen Sehsystems. Das Ziel der CNNs ist es, das visuelle System des Menschen nachzuahmen.

Es können verschiedene Neurone auf unterschiedliche lokale Bereiche des Eingabebildes reagieren. Auf die Faltungsschicht folgt eine Aktivierungsschicht, um komplexere Eigenschaften des Eingangssignals zu bearbeiten. Poolingsschichten werden verwendet, um überflüssige Informationen zu verwerfen. Schließlich werden mehrere Schichten zu einer zusammengefasst, die letztendlich das Klassifizierungsergebnis liefert (Anthimopoulos, Christodoulidis et al. 2016, Mocanu, Mocanu et al. 2018).

Ein künstliches Neuronales Netz (KNN) ist aus künstlichen Neuronen zusammengesetzt. Ziel des KNNs ist es das menschliche Neuronale Netz zu imitieren, um komplexe Aufgaben besser lösen zu können (Jain, Jianchang et al. 1996). Das Interesse an die Anwendung setzte bereits in den 1940er Jahren ein. Im Jahr 1943 wurden die ersten Beispiele von Warren McCulloch, einem Neurowissenschaftler und dem Mathematiker Walter Pitts eingeführt (McCulloch and Pitts 1943). Künstliche Neurone können auf verschiedenster Weise miteinander verbunden sein und eine unterschiedliche Anzahl an Neuronenschichten aufweisen. Einschichtige Netze stellen die einfachste Form dar. Sind mehr als eine Schicht vorhanden, spricht man von einem mehrschichtigen Netz. Die

letzte Schicht des Netzes wird als Ausgabeschicht (Output Layer) und die davorliegenden Schichten als verdeckte Schichten (Hidden Layers) bezeichnet (Zhang, Eddy Patuwo et al. 1998). Besitzt ein Netz eine Eingabeschicht (Input Layer), mehrere verborgene Schichten und eine Ausgabeschicht, wird von einem tiefgehenden Netz (Deep Neural Network, DNN) gesprochen (Abbildung 5). Wie viele Ebenen notwendig sind, um von einem DNN zu sprechen, ist nicht fest definiert. Wird ein solches DNN mit Daten angelern spricht man von tiefgehendem Lernen (Deep Learning) (Schmidhuber 2015).

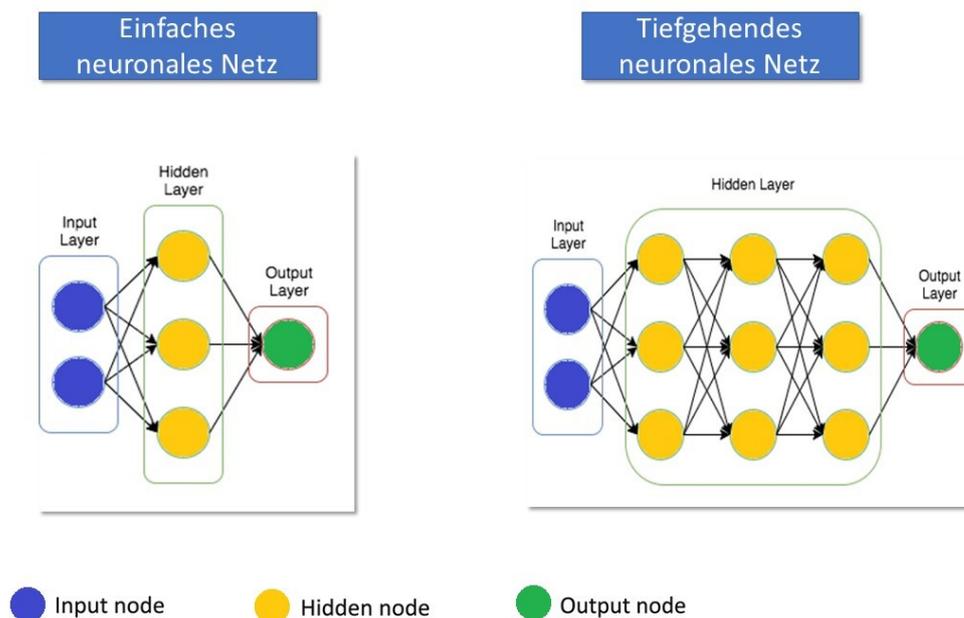


Abbildung 5: einfaches vs. tiefgehendes Neuronales Netz

Ein DNN kann von Grund aus neu angelern werden (Training from Scratch) oder es können vorbestehende, bereits angelernete Netzwerke neu angelern werden (Transfer Learning). Abbildung 6 gibt den wesentlichen Unterschied zwischen diesen beiden Anlernmethoden wieder.

Durch ihre hohe Genauigkeit haben im Bereich der Bilderkennung und Bildklassifikation CNNs eine besondere Stellung. Beispiele für bereits vorangelernete CNNs sind Inception (He, Zhang et al. 2016), ResNet (He, Zhang et al. 2016) und DenseNet (Huang, Liu et al. 2017).

Zum Anlernen eines Neuronales Netzes sind viele validierte Daten essentiell. Die zur Verfügung stehenden Daten sollten in drei Gruppen aufgeteilt werden: Lerndatensatz,

Validierungsdatensatz und Prüfdatensatz. Abbildung 7 veranschaulicht das Aufteilungsprinzip der Daten.

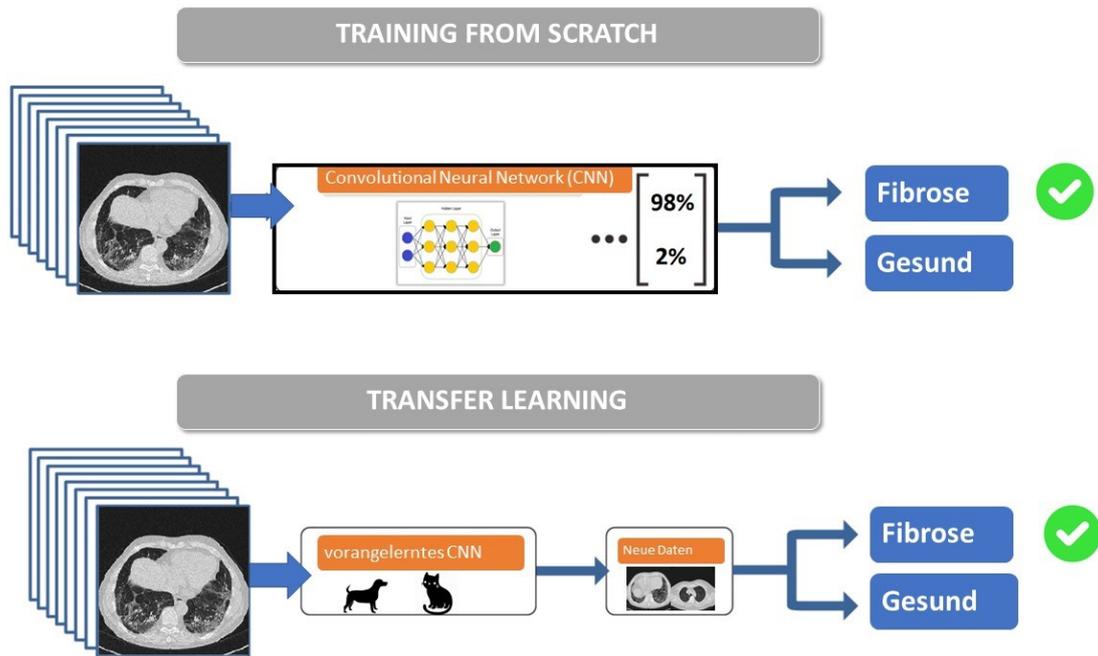


Abbildung 6: Lernansätze – Training from Scratch vs. Transfer Learning. Bei Transfer Learning werden vorangelernte Netzwerke, die beispielsweise zwischen Hunden und Katzen differenzieren können mit neuen Daten neu angelehrt.

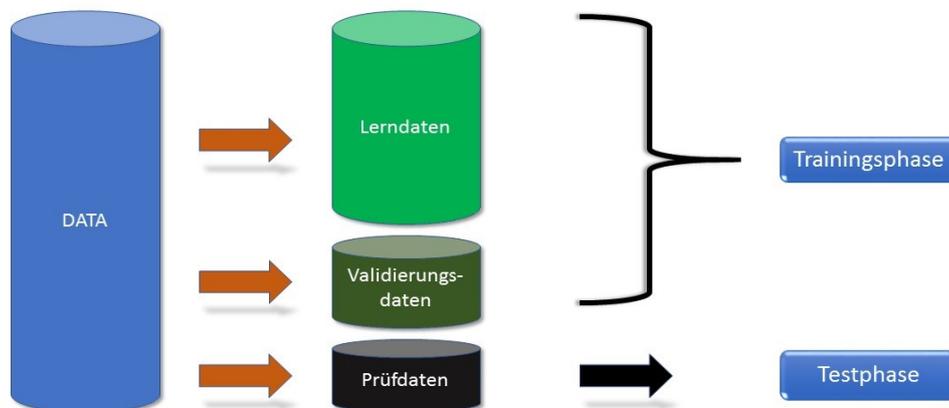


Abbildung 7: Aufteilungsprinzip der Daten

Zielsetzung der Arbeit

Die radiologische Bildgebung wird immer präziser und findet dadurch häufiger Anwendung im klinischen Alltag. Eine zuverlässige Diagnose ist in der heutigen modernen Welt ohne Bildgebung nahezu undenkbar. Die Radiologen werden daher immer mehr einer höheren Arbeitsbelastung ausgesetzt. Eine computergestützte Hilfestellung bei Befundung radiologischer Bildgebung wäre eine große Bereicherung im klinischen Alltag.

Das Ziel dieser Arbeit ist es, mit relativ wenig vorhandenen Bilddaten, Deep Learning Algorithmen zur vollautomatischen Differenzierung zwischen dem Lungenparenchym „Gesund“ und „Fibrose“ zu optimieren, wobei zertifizierte CT-Schnittbilder die Klassifizierungsgrundlage bilden. In Zukunft können diese Methoden eingesetzt werden, um Deep Learning Modelle für seltenere Erkrankungen mit wenig vorhandenen Daten zu entwickeln.

Bei in Deutsch verfassten Dissertationen sind Anglizismen zu vermeiden. Auf dem Gebiet der künstlichen Intelligenz werden jedoch sehr viele Fachbegriffe, die in der englischen Sprache eingeführt wurden, in der deutschen Sprache ohne eine Übersetzung verwendet. In diesen Fällen wurden auch in dieser Arbeit die englischen Fachtermini benutzt.

Auf Beschreibung technischer Details, die nicht zum Verständnis dieser Arbeit beitragen, wurde verzichtet. Alle Abbildungen und Tabellen wurden im Rahmen der Dissertation selbst erstellt.

Material und Methoden

Patienten

Ein positives Votum der Ethikkommission des Fachbereichs Medizin lag vor Beginn der Untersuchung vor (Aktenzeichen Projekt 222/16). Aus dem Zeitraum von 2011 und 2016 wurden aus dem Pool des Bilddatenarchivs des Universitätsklinikums Gießen 52 HRCT-Datensätze mit unauffälligen Thoraxbefunden (23 männlich, mittleres Alter 50,6 Jahre, Bereich 22-79 Jahre; 29 weiblich, mittleres Alter 46,9 Jahre, Bereich 19-80 Jahre) und 52 Patienten mit fibrotisch erkranktem Lungenparenchym anhand entsprechender Befunde herausgesucht (53 männlich, mittleres Alter 69,6, Bereich 51-86 Jahre; 17 weiblich, mittleres Alter 66,7 Jahre, Bereich 52-76 Jahre). Daten von Patienten mit starken Bewegungsartefakten und erheblichen Mischdiagnosen wurden ausgeschlossen. Die Diagnosen wurden nochmals durch einen Radiologen mit 10-jähriger Berufserfahrung und stichprobenartig von einem Radiologen mit 25 Jahren Berufserfahrung kontrolliert. Unauffällige, normale Thoraxbefunde zeichnen sich durch die Abwesenheit pulmonaler Pathologien oder Abnormalitäten aus. Als solche gelten CT-morphologisch erkennbare bronchiale, pulmonale, vaskuläre oder pleurale Strukturen. Zur Fibrosedefinition wurden die Klassifikationskriterien der "American Thoracic Society" und der "European Thoracic Society" zu „idiopathic interstitial pneumonia“ (IIP) herangezogen. HRCT- Muster mit dem Bild einer „non-specific interstitial pneumonia“ (NSIP) und „usual interstitial pneumonia“ (UIP), wurden als Fibrose eingeschlossen. Als Mischdiagnosen wurden Datensätze ausgeschlossen, bei denen neben der bekannten Lungenfibrose weitere Erkrankungen (Pneumonie, Bronchialkarzinome/Metastasen, Emphysem, Atelektase) vorlagen.

CT-Bildgebung

Die CT Untersuchungen wurden mit einem 64- oder einem 40- Zeilen CT Gerät (Somatom Definition Dual Source oder Somatom Definition AS; Siemens Medical Systems, Erlangen, Deutschland) durchgeführt. Das folgende Untersuchungsprotokoll wurde angewendet: Der Patient wird in Rückenlage positioniert, Kopf voraus; vollständige Lungenabbildung vom Apex bis zum Diaphragma nach tiefem Einatmen und

anschließendem Luft anhalten; 120 kV; die Spannung wurde nach Patientengewicht variiert: < 75 kg: 25 mAs; > 75 kg: 45 mAs bei dem 64 MSCT und 50 mAs bei dem 40 MSCT; Pitch 1.2; Schichtdicke 1 mm; Bildrekonstruktion mit Kernel B50f in eine 512×512 Bildmatrix.

TensorFlow

Auf einem Notebook mit dem Betriebssystem Windows 8 wurde zunächst eine Linux-Distribution installiert (Ubuntu 17.04). Nach erfolgreichem Installieren wurden durch folgende Befehle die Programmbibliothek TensorFlow, andere notwendige Programmbibliotheken und Python eingerichtet:

```
1. $ sudo apt-get install python-pip python-dev python-numpy python-wheel
2. $ sudo pip install --upgrade pip
3. $ sudo pip install tensorflow #CPU-Version
4. $ sudo apt-get install git
5. $ sudo git clone https://github.com/tensorflow/tensorflow
6. $ sudo add-apt-repository ppa:webupd8team/java && sudo apt-get update
7. $ sudo apt-get install oracle-java8-installer
8. $ curl $ https://bazel.build/bazel-release.pub.gpg | sudo apt-key add -
9. $ sudo apt-get update
10. $ echo "deb [arch=amd64] http://storage.googleapis.com/bazel-
    apt stable jdk1.8" | sudo tee /etc/apt/sources.list.d/bazel.list
11. $ sudo apt-get update && sudo apt-get install bazel
12. $ sudo pip install --upgrade \https://storage.googleapis.com/tensor-
    flow/linux/cpu/protobuf-3.1.0-cp27-none-linux_x86_64.whl
13. $ sudo pip install --upgrade \https://storage.googleapis.com/tensor-
    flow/linux/cpu/protobuf-3.1.0-cp27-none-linux_x86_64.whl
14. $ cd tensorflow
15. $ sudo ./configure
16. $ sudo bazel build --config opt tensorflow/examples/image_retraining:retrain
17. $ sudo bazel build tensorflow/examples/image_retraining:retrain --local_re-
    sources 2048,2.0,1.0
```

Im Dezember 2016 wurde das vorangelernte Netzwerk Inception v3 unter folgendem Link heruntergeladen:

http://download.tensorflow.org/models/inception_v3_2016_08_28.tar.gz.

Arbeitsablauf

Zunächst wurden 104 Patientendatensätze aus dem Bildarchiv des Universitätsklinikums Gießen anhand entsprechender Befundeinträge herausgesucht und die Richtigkeit der Diagnosen von Radiologen nochmals überprüft. Nach Anonymisieren der Daten wurden alle Schnittbilder patientenspezifisch im Lungenfenster als JPEG abgespeichert. Danach wurden alle Patientenbilder auf eine externe Festplatte kopiert. Manuell wurden anschließend alle Schnittbilder gelöscht, auf denen kein Lungengewebe abgebildet war.

Zum Anlernen Neuronaler Netze ist die Datenstruktur essentiell. Es sollte dabei sichergestellt werden, dass Lehr- und Prüfdaten nicht in verwechselbaren Subordnern gespeichert wurden. Um die richtige Datenstruktur zu erhalten wurden zwei Python Programme geschrieben, wobei das erste alle Schnittbilder mit einer Laufnummer versieht und somit unterschiedliche Benennungen sicherstellt (Code 1). Das zweite Programm kopiert alle Schnittbilder aus den Ursprungsordnern und fügt sie in für TensorFlow eindeutig benannte Ordner (Gesund und Fibrose) ein (Code 2). Abbildung 8 zeigt die notwendige Datenstruktur für TensorFlow auf.

Code 1

```
1. import os
2.
3.
4. path = ""
5. count = 1
6.
7. for root, dirs, files in os.walk(path):
8.     for i in files:
9.         os.re-
            name(os.path.join(root, i), os.path.join(root, str(count) + ".jpg"))
10.        count += 1
```

Code 2

```
1. import os
2. import shutil
3.
4. for root, dirs, files in os.walk(''): # starting directory
5.     for file in files:
6.         path_file = os.path.join(root,file)
7.         shutil.copy2(path_file,') # destination dir
```

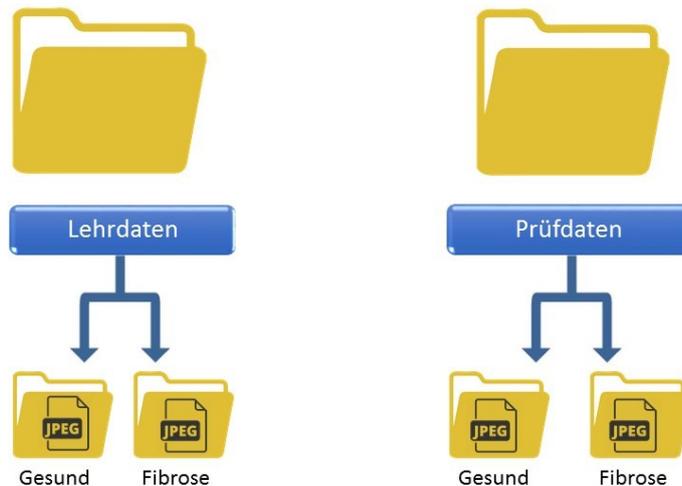


Abbildung 8: Datenstruktur für TensorFlow

Bilder im Ordner „Lehrdaten“ überschneiden sich nicht mit Bildern im Ordner „Prüfdaten“. Der Lernvorgang wurde mit folgendem Terminalbefehl gestartet:

```
1. $ sudo Downloads/tensorflow/bazel-bin/tensorflow/examples/image_retraining/re-train --image_dir /home/lynx/Desktop/Lehrdaten
```

Für die Anlernphase wurden folgende Parameter verwendet:

- Epoche (Anzahl der Trainingsdurchläufe der Input-Daten): 8000
- Anlernrate: 0.01
- Validierungsdatensatz: 10%
- Trainingsbatchgröße: 100
- Validierungsbatchgröße: 100 (Datenmenge, die nach jeder Lerneinheit zur internen Validierung herangezogen wird)

Nach erfolgreichem Anlernen wurde eine ca. 90 MB große Graph-Datei und eine Textdatei mit den Klassifikationslabeln (Gesund und Fibrose) abgespeichert. Durch einen weiteren Terminalbefehl wurde ein weiteres Python Programm gestartet (Code 3). Dieses wertete alle Prüfbilder aus und speicherte die Klassifikationsergebnisse und die dazu gehörigen Zuordnungswahrscheinlichkeiten in eine Textdatei ab. Die Anzahl richtig bzw. falsch zugeordneter Bilder, Spezifität und Sensitivität wurden abschließend mit einem Tabellenkalkulationsprogramm berechnet (Microsoft® Excel® Version 2013).

Code 3

```
1. import numpy as np
2. import tensorflow as tf
3. import argparse
4. from PIL import Image
5. import sys
6. import os
7.
8. # construct the argument parse and parse the arguments
9. ap = argparse.ArgumentParser()
10. ap.add_argument("-i", "--image", required = True,
11.     help = "path to input dataset of images")
12. args = vars(ap.parse_args())
13.
14. imagePath = os.path.abspath(args["image"])
15. modelFullPath = '/home/lynx/Desktop/output_graph.pb'
16. labelsFullPath = '/home/lynx/Desktop/output_labels.txt'
17.
18.
19. def create_graph():
20.     """Creates a graph from saved GraphDef file and returns a saver."""
21.     # Creates graph from saved graph_def.pb.
22.     with tf.gfile.FastGFile(modelFullPath, 'rb') as f:
23.         graph_def = tf.GraphDef()
24.         graph_def.ParseFromString(f.read())
25.
26.     del(graph_def.node[1].attr["dct_method"])
27.     _ = tf.import_graph_def(graph_def, name='')
28.
29.
30. def run_inference_on_image():
31.     answer = None
32.
33.     if not tf.gfile.Exists(imagePath):
34.         tf.logging.fatal('File does not exist %s', imagePath)
35.         return answer
36.
37.     image_data = tf.gfile.FastGFile(imagePath, 'rb').read()
38.
39.     # Creates graph from saved GraphDef.
40.     create_graph()
41.
42.     with tf.Session() as sess:
43.
44.         softmax_tensor = sess.graph.get_tensor_by_name('final_result:0')
45.         predictions = sess.run(softmax_tensor,
46.             {'DecodeJpeg/contents:0': image_data})
47.         predictions = np.squeeze(predictions)
48.
49.         top_k = predictions.argsort()[-5:][::-1] # Getting predictions
50.         top_k1 = predictions.argsort()[-1:] # Getting Diagnosis
51.         f = open(labelsFullPath, 'rb')
52.         lines = f.readlines()
53.         labels = [str(w).replace("\n", "") for w in lines]
54.         for node_id in top_k:
55.             human_string = labels[node_id]
56.             score = predictions[node_id]
57.             print('%s (score = %.5f)' % (human_string, score))
58.
59.         for node_id in top_k1:
60.             human_string = labels[node_id]
```

```

61.         score = predictions[node_id]
62.
63.         print 'Diagnose: %s' %human_string
64.
65.         answer = labels[top_k[0]]
66.         return answer
67.
68.
69. if __name__ == '__main__':
70.     run_inference_on_image()

```

Anzahl der Bilddaten

Es wurden zwei Varianten mit unterschiedlicher Anzahl der Lehr- und Prüfdaten ausgetestet. (Variante 1; Variante 2)

Variante 1 – Bildspezifisch

Ein Python-Programm wurde erstellt (Code 4), das zufällig Bilder aus den Ordnern „Gesund“ bzw. „Fibrose“ auswählt und in die entsprechenden Ordner „Lehrdaten_Gesund“ (1351Bilder), „Lehrdaten_Fibrose“ (942 Bilder), „Prüfdaten_Gesund“ (1169 Bilder) und „Prüfdaten_Fibrose“ (1057 Bilder) kopiert. Auch die Anzahl der Bilder der einzelnen Gruppen wurde von einem Zufallszahlengenerator erzeugt und wählte Bilder beliebiger Patienten aus.

Code 4

```

1. import os
2. import shutil
3. import random
4.
5. root_dir = ''
6. output_dir = ''
7. r = 1
8.
9. for root, dirs, files in os.walk(root_dir):
10.     number_of_files = len(os.listdir(root))
11.     if number_of_files > r:
12.         files_move = random.randint(1,5000) # amount to copy
13.         for i in xrange(files_move):
14.             chosen_one = random.choice(os.listdir(root))
15.             file_in_track = root
16.             file_to_move = file_in_track + '/' + chosen_one
17.             if os.path.isfile(file_to_move) == True:
18.                 shutil.move(file_to_move,output_dir)
19.                 print file_to_move
20.     else:
21.         for i in xrange(len(files)):
22.             track_list = root

```

```

23.         file_in_track = files[i]
24.         file_to_move = track_list + '/' + file_in_track
25.         if os.path.isfile(file_to_move) == True:
26.             shutil.move(file_to_move,output_dir)
27.             print file_to_move
28. print 'Done !'

```

Variante 2 – Patientenspezifisch

Es wurden jeweils 26 zufällig gewählte Patienten diagnosespezifisch in die Ordner „Lehrdaten_Gesund“ (5622 Schnittbilder), „Lehrdaten_Fibrose“ (3323 Schnittbilder), „Prüfdaten_Gesund“ (5407 Schnittbilder) und „Prüfdaten_Fibrose“ (2919 Schnittbilder) kopiert (modifizierter Code 3).

PyTorch

Für die Berechnungen unter PyTorch wurden die Lehrdaten (26 Patienten) nochmals in Lerndaten (23 Patienten) und Validierungsdaten (3 Patienten) unterteilt. Abbildung 9 gibt die dabei entstandene neue Datenstruktur wieder.



Abbildung 9: Datenstruktur für PyTorch

Durch den folgenden Terminal-Befehl wurde die Programm-bibliothek PyTorch eingerichtet:

```

1. pip3 install torch torchvision

```

Anschließend wurden einige PyTorch Skripte heruntergeladen (<https://github.com/pytorch/tutorials>, Zugriff am 08.02.2019 um 19:51 Uhr).

Anhand dieser Skripte und den bereits eingestellten Parametern wurden Inception v3, ResNet50 und DenseNet121 neu angelern. Die neu angelernen Modelle wurden im .pt Format abgespeichert.

Ein weiteres Python-Skript (Jupyter Notebook) wurde zur automatischen Klassifizierung der Prüfdaten und zum Speichern der Ergebnisse in eine Textdatei erstellt (Code 5). Die Anzahl richtig bzw. falsch zugeordneter Bilder, Spezifität und Sensitivität wurden abschließend mit einem Tabellenkalkulationsprogramm berechnet (Microsoft® Excel® Version 2013).

Code 5

```
1. from __future__ import print_function, division
2. %matplotlib inline
3. import torch
4. import torch.nn as nn
5. import torch.optim as optim
6. from torch.optim import lr_scheduler
7. import numpy as np
8. import torchvision
9. from torchvision import datasets, models, transforms
10. import matplotlib.pyplot as plt
11. import time
12. import os
13. import copy
14. from PIL import Image
15. from torch.autograd import Variable
16. from torch.nn import functional as F
17. from torch import topk
18. from matplotlib.pyplot import imshow
19. import matplotlib.pyplot as plt
20.
21. plt.ion() # interactive mode
22.
23. import os
24. os.environ["CUDA_DEVICE_ORDER"]="PCI_BUS_ID"
25. os.environ["CUDA_VISIBLE_DEVICES"] = '0'
26. #os.environ['USE_CUDA'] = 'True'
27.
28. data_dir = '/home/qmohammad/Lerndaten_pytorch/Testdaten/0/'
29. folder_name = 'Gesund'
30. directory = os.path.join(data_dir, folder_name)
31. path, dirs, files = next(os.walk(directory))
32. file_count = len(files)
33. print (file_count)
34.
35. data_transforms = {
36.     folder_name: transforms.Compose([
37.         transforms.Grayscale(num_output_channels=3),
38.         transforms.Resize(224,224),
39.         #transforms.CenterCrop(224),
40.         transforms.ToTensor(),
41.         transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
42.     ]),
43. }
44.
```

```

45.
46. image_datasets = {x: datasets.ImageFolder(os.path.join(data_dir),
47.                                     data_transforms[x])
48.                   for x in [folder_name]}
49. dataloaders = {x: torch.utils.data.DataLoader(image_da-
50.         tasetypes[x], batch_size=4,
51.                                     shuffle=True, num_workers=4)
52.               for x in [folder_name]}
53. dataset_sizes = {x: len(image_datasets[x]) for x in [folder_name]}
54. class_names = [0,1]
55.
56. #device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
57. device = torch.device(0)
58.
59. def imshow(inp, title=None):
60.     """Imshow for Tensor."""
61.     inp = inp.numpy().transpose((1, 2, 0))
62.     mean = np.array([0.485, 0.456, 0.406])
63.     std = np.array([0.229, 0.224, 0.225])
64.     inp = std * inp + mean
65.     inp = np.clip(inp, 0, 1)
66.     plt.imshow(inp)
67.     if title is not None:
68.         plt.title(title)
69.     plt.pause(0.001) # pause a bit so that plots are updated
70.
71.
72. class SaveFeatures():
73.     features=None
74.     def __init__(self, m): self.hook = m.register_forward_hook(self.hook_fn)
75.     def hook_fn(self, module, input, output): self.features = ((out-
76.         put.cpu()).data).numpy()
77.     def remove(self): self.hook.remove()
78.
79. final_layer = model._modules.get('features') #Densenet
80. #final_layer = model._modules.get('Mixed_7b') #inception
81. #final_layer = model._modules.get('layer4') #ResNet
82.
82. def visualize_model(model, num_images=file_count):
83.     was_training = model.training
84.     model.eval()
85.     images_so_far = 0
86.     fig = plt.figure(figsize=(50,50))
87.     for filename in os.listdir(directory):
88.         if filename.endswith(".jpg"):
89.             image = Image.open(os.path.join(directory, filename))
90.             normalize = transforms.Normalize(
91.                 mean=[0.485, 0.456, 0.406],
92.                 std=[0.229, 0.224, 0.225]
93.             )
94.             preprocess = transforms.Compose([
95.                 transforms.Resize((224,224)),
96.                 transforms.Grayscale(num_output_channels=3),
97.                 transforms.ToTensor(),
98.                 normalize
99.             ])
100.            display_transform = transforms.Compose([
101.                transforms.Resize((224,224))]
102.            )
103.            tensor = preprocess(image)
104.            prediction_var = Variable((tensor.unsqueeze(0)).cuda(), re-
105.                quires_grad=True)
106.            activated_features = SaveFeatures(final_layer)
107.            prediction = model(prediction_var)
108.            pred_probabilities = F.softmax(prediction).data.squeeze()
109.            activated_features.remove()

```

```

108.         f = open('/home/qmohammad/Lerndaten_all_pytorch/test_Ge-
sund_Densenet.txt', 'a') #Inception?Resnet?Densenet?
109.         print(filename, topk(pred_probabilities,1), file = f)
110.         continue
111.     else:
112.         continue
113.
114.     #Resnet
115.     #model = models.resnext50_32x4d(pretrained=True)
116.     #num_ftrs = model.fc.in_features
117.     #model.fc = nn.Linear(num_ftrs, 2)
118.     #model.load_state_dict(torch.load('/home/qmoham-
mad/Lerndaten_all_pytorch/GF_ResNet.pt'))
119.     #model.eval()
120.     #model.cuda(device)
121.
122.
123.     #Inception
124.     #model = models.inception_v3(pretrained=True)
125.     #model.fc = nn.Linear(2048, 2)
126.     #model.aux_logits = False
127.     #model.load_state_dict(torch.load('/home/qmoham-
mad/Lerndaten_all_pytorch/GF_Inception.pt'))
128.     #model.eval()
129.     #model.to(device)
130.
131.
132.     #DenseNet
133.     model = models.densenet121(pretrained=True)
134.     model.classifier = nn.Linear(1024, 2)
135.     model.load_state_dict(torch.load('/home/qmoham-
mad/Lerndaten_all_pytorch/GF_Densenet.pt'))
136.     model.eval()
137.     model.to(device)
138.
139.     visualize_model(model)

```

Um die neu angelernten Modelle bei ihrer Klassifizierungsentscheidung transparenter darzustellen, wurden für einige gewählte Bilder Class-Activation-Maps (CAM) erstellt.

Class-Activation-Maps (CAMs)

Class-Activation-Maps (Heat Maps) sind eine einfache Methode, um die Bildregionen hervorzuheben, anhand welcher das neuronale Netz die spezifische Gruppenzuordnung veranlasst. Das verwendete neuronale Netz sollte hierfür die Schicht „Global Average Pooling (GAP)“ besitzen. Abbildung 10 gibt einen Überblick über die Funktionsweise der CAMs. Das Eingabebild wird abhängig vom verwendeten neuronalen Netz in unterschiedliche Anzahl von Pixelpunkten unterteilt. Beim Verwenden von DenseNet121 sind dies beispielsweise 7x7 Pixelpunkte. Abhängig von der veranlassten Klassifikation wird für jeden Pixelpunkt nach der GAP-Schicht eine Gewichtung (w) bestimmt. Die Gewichtung der einzelnen Pixelpunkte ist umso höher, je bedeutsamer diese Bildregion für die

veranlasste Klassifikation ist. Höhere Gewichtungszahlen werden mit warmen Farben dargestellt. Pixelpunkte, die nicht oder wenig zur Klassifikation beitragen, werden mit kalten Farben abgebildet. Das somit entstandene Bild wird anschließend mit dem Eingabebild überlagert. Heat Maps sind essentiell, um Klassifizierungen nachzuvollziehen und lassen die Entscheidungsfindung neuronaler Netze transparenter darstellen. Für das Erstellen der CAMs wurde das folgende Skript (Code 6) verwendet.

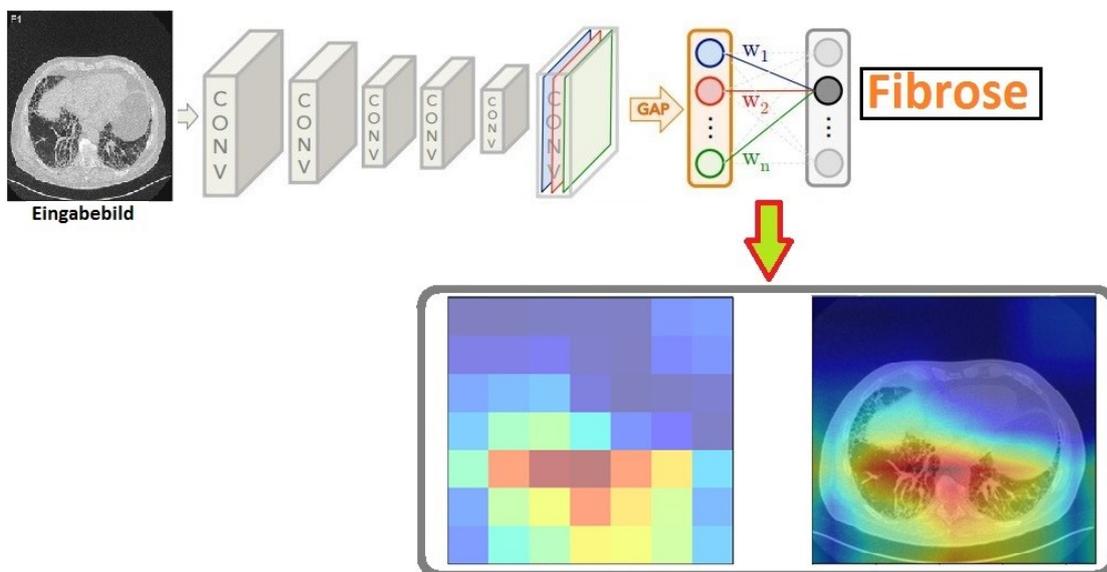


Abbildung 10: Beispiel einer Class-Activation-Map

Code 6

```

1. class SaveFeatures():
2.     features=None
3.     def __init__(self, m): self.hook = m.register_forward_hook(self.hook_fn)
4.     def hook_fn(self, module, input, output): self.features = ((out-
5.         put.cpu()).data).numpy()
6.     def remove(self): self.hook.remove()
7.
8. #final_layer
9. #final_layer = model._modules.get('Mixed_7c') #inception
10. #final_layer = model._modules.get('layer4') #ResNet
11. #final_layer = model._modules.get('features') #Densenet
12.
13. activated_features = SaveFeatures(final_layer)
14.
15. prediction = model(prediction_var)
16. pred_probabilities = F.softmax(prediction).data.squeeze()
17. activated_features.remove()
18.
19. print(topk(pred_probabilities,1))
20.
21. def getCAM(feature_conv, weight_fc, class_idx):
22.     _, nc, h, w = feature_conv.shape
23.     cam = weight_fc[class_idx].dot(feature_conv.reshape((nc, h*w)))
24.     cam = cam.reshape(h, w)
25.     cam = cam - np.min(cam)

```

```

25.     cam_img = cam / np.max(cam)
26.     return [cam_img]
27.
28. weight_softmax_params = list(model.parameters())
29. weight_softmax = np.squeeze(weight_softmax_params[-2].cpu().data.numpy())
30.
31. #weight_softmax_params = list(model._modules.get('fc').parameters())
32. #weight_softmax = np.squeeze(weight_softmax_params[0].cpu().data.numpy())
33.
34. weight_softmax_params
35. class_idx = topk(pred_probabilities,1)[1].int()
36. overlay = getCAM(activated_features.features, weight_softmax, class_idx )
37. imshow(overlay[0], alpha=0.5, cmap='jet')
38. plt.savefig(filename + '_heat_map.png', dpi=1200)

```

Genutzte Computerhardware

Alle TensorFlow-Berechnungen (Anlernphase, Klassifizierung, Python-Programm-Erstellungen und Programm-Ausführungen) wurden auf einem Standard-Notebook unter dem Betriebssystem Ubuntu mit einem 2,1 (Quad-Core) GHz-Prozessor und 3,8 GByte RAM durchgeführt (CPU-Berechnungen).

Alle PyTorch Kalkulationen (Anlernphase, Klassifizierungen, Python-Programm-Ausführungen, CAM) wurden auf einem GPU-fähigen Rechner mit dem Betriebssystem Ubuntu durchgeführt.

Ergebnisse

Die neu angelernen Neuronalen Netze konnten erfolgreich eingesetzt werden und zeigten, abhängig von verwendeten Lerndaten und Netzwerkstruktur unterschiedliche Ergebnisse in Sensitivität und Spezifität. Für die zufällig gewählte, bildspezifische Aufteilung der Daten (Variante 1, TensorFlow) zeigte sich eine Sensitivität von 96,9 % und eine Spezifität von 95,4 %. Nach Erhöhung der Anzahl und nach patientenspezifischer Aufteilung der Daten erhöhten sich Sensitivität und Spezifität auf > 99%, (Variante 2). Die neu angelernen Modell-Varianten 3 bis 5 unter PyTorch (DenseNet121, Inception v3 und ResNet50) zeigten eine Spezifität von 100 %. Inception v3 und DenseNet121 klassifizierten alle fibrotischen Bilder als „Fibrose“ ein und zeigten somit auch eine Sensitivität von 100 %. ResNet50 ordnete fälschlicherweise 2 fibrotische Bilder als „Gesund“ ein (Sensitivität 99,93 %).

Abbildungen 11-23 zeigen Beispiele richtig bzw. falsch klassifizierter Bilder mit ihren zugehörigen Zuordnungswahrscheinlichkeiten (N = Gesund, F = Fibrose) in Abhängigkeit genutzter Programmbibliothek und des verwendeten Neuronalen Netzes. Abbildungen 24-26 fassen die Lern- und Validierungsgenauigkeit während der Anlernphase unter PyTorch zusammen. In den Graphiken wurde dabei jeweils die erzielte Genauigkeit (Accuracy) gegen die Zyklusanzahl der Berechnungen (Epoch) aufgetragen.

Tabellen 1-5 geben die richtig und falsch zugeordneten Bilder in Abhängigkeit verwendeter Programmbibliothek (TensorFlow vs. PyTorch) und des verwendeten Neuronalen Netzes mit unterschiedlichen Bereichen der Zuordnungswahrscheinlichkeiten S im Detail wieder.

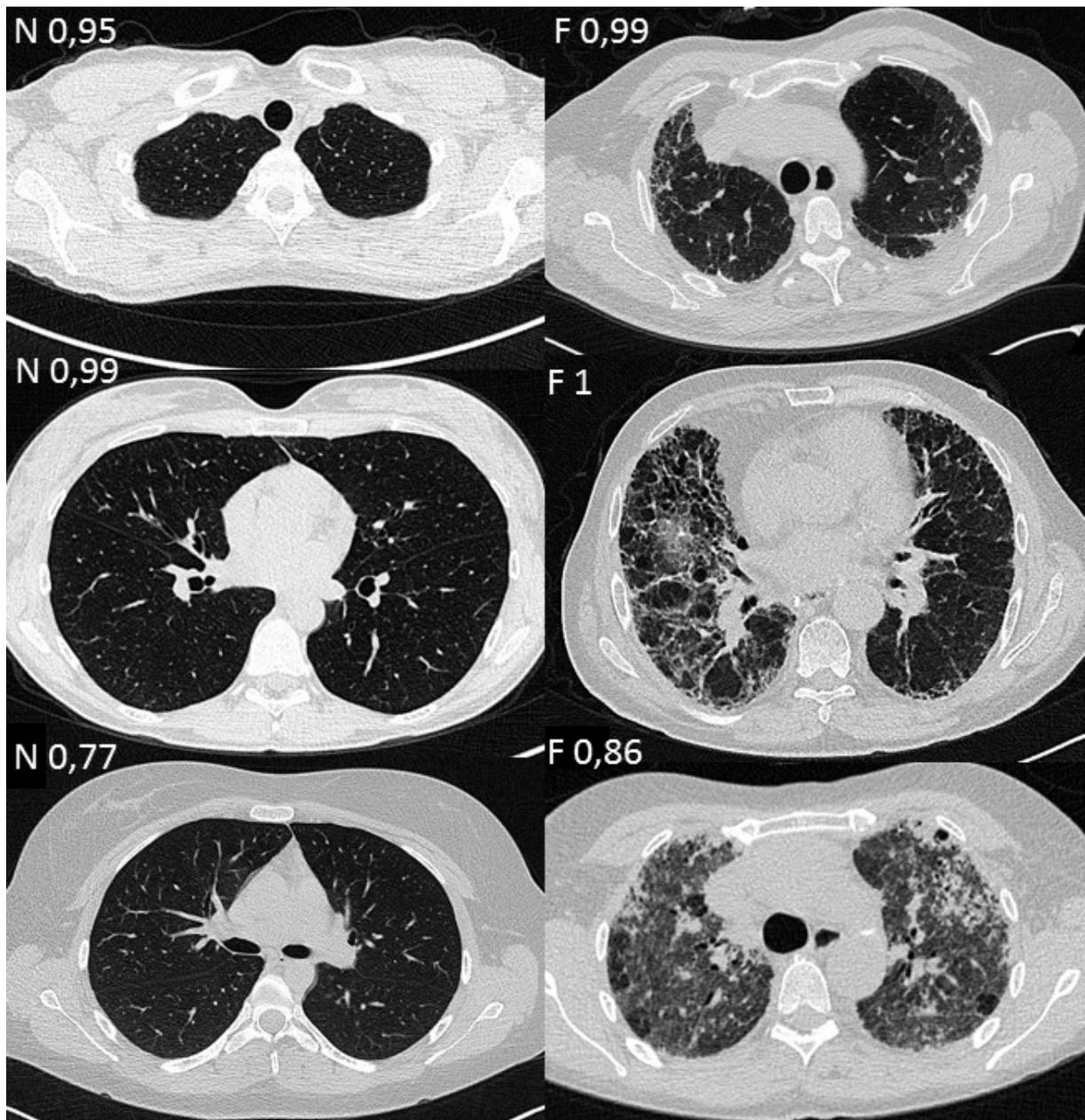


Abbildung 11: richtig klassifiziert (TensorFlow, Inception v3, Variante 1)

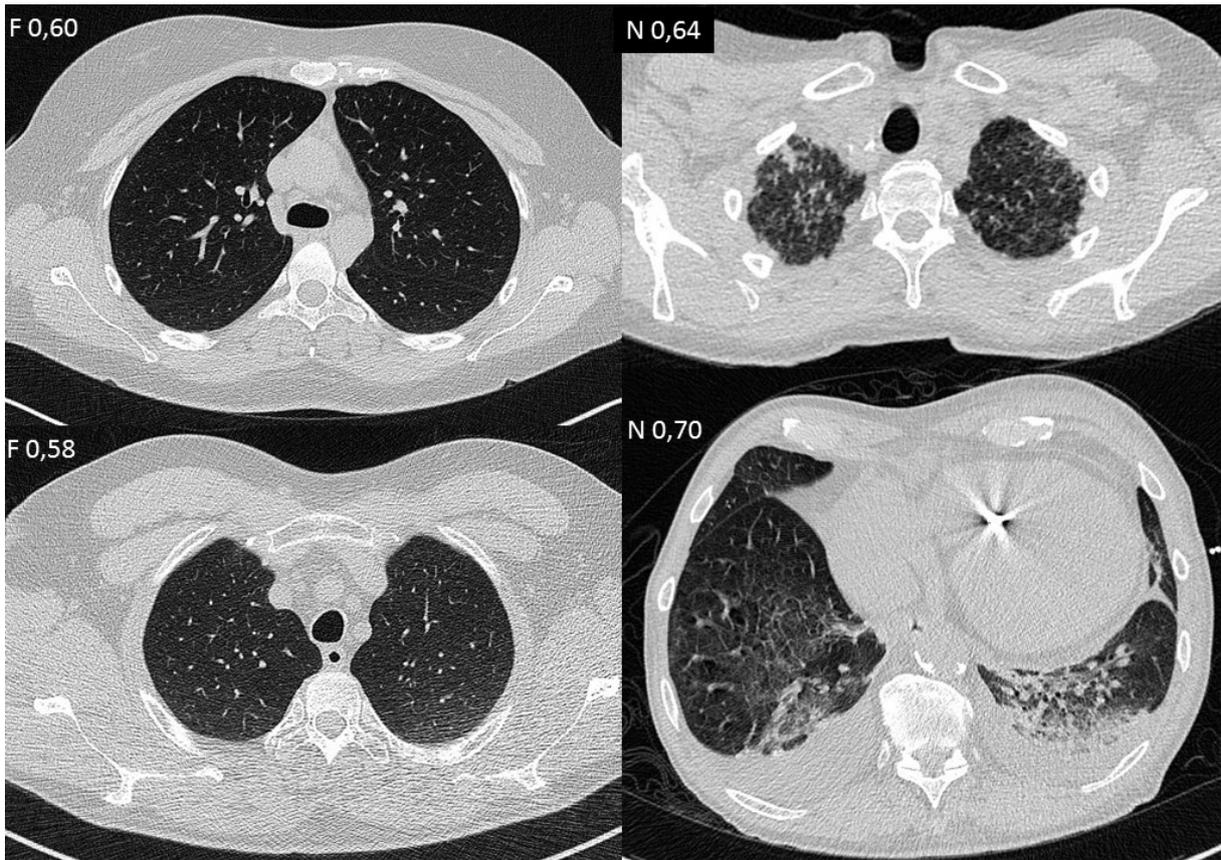


Abbildung 12: falsch klassifiziert (TensorFlow, Inception v3, Variante 1)

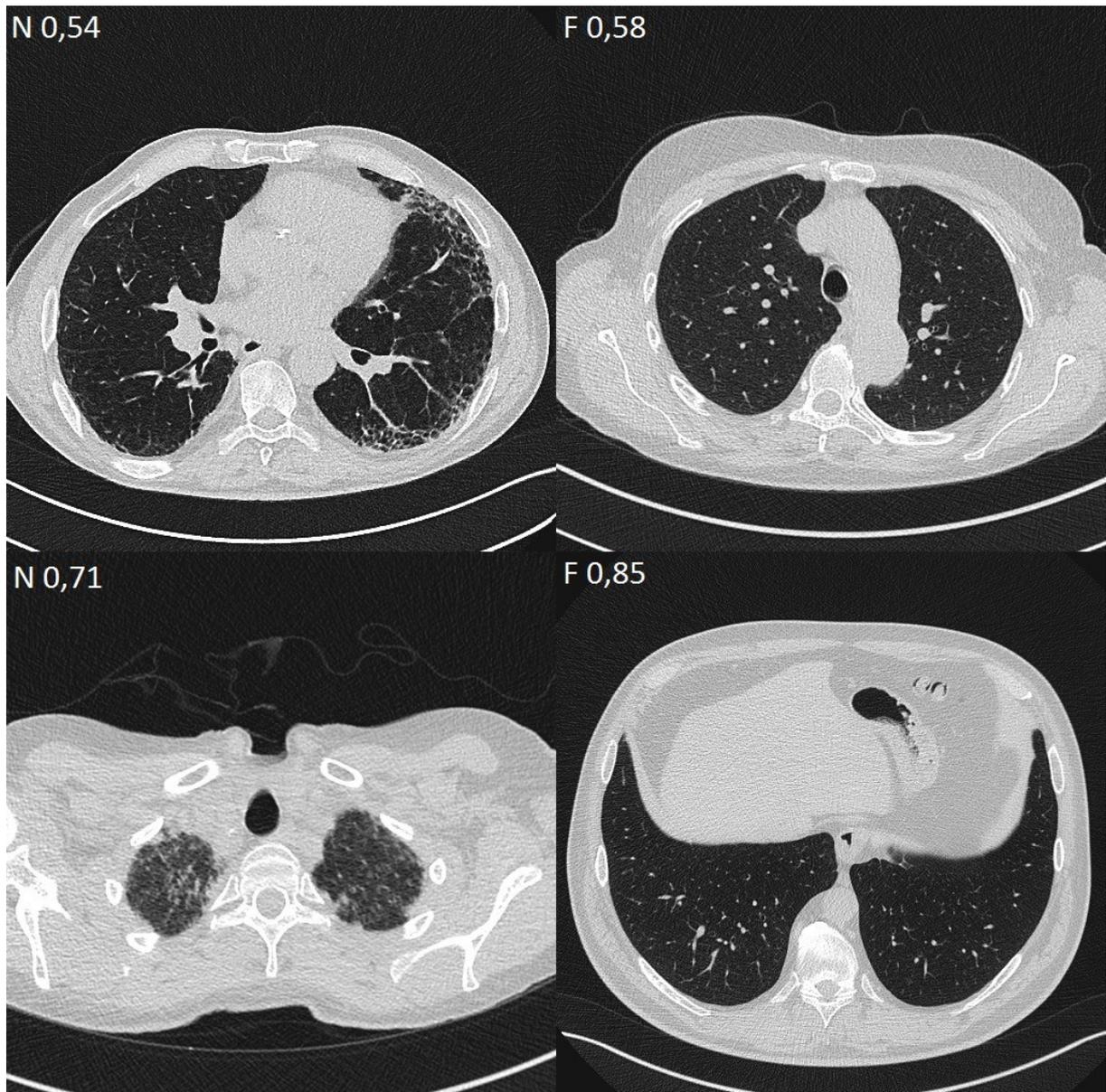


Abbildung 13: falsch klassifiziert (TensorFlow, Inception v3, Variante 2)

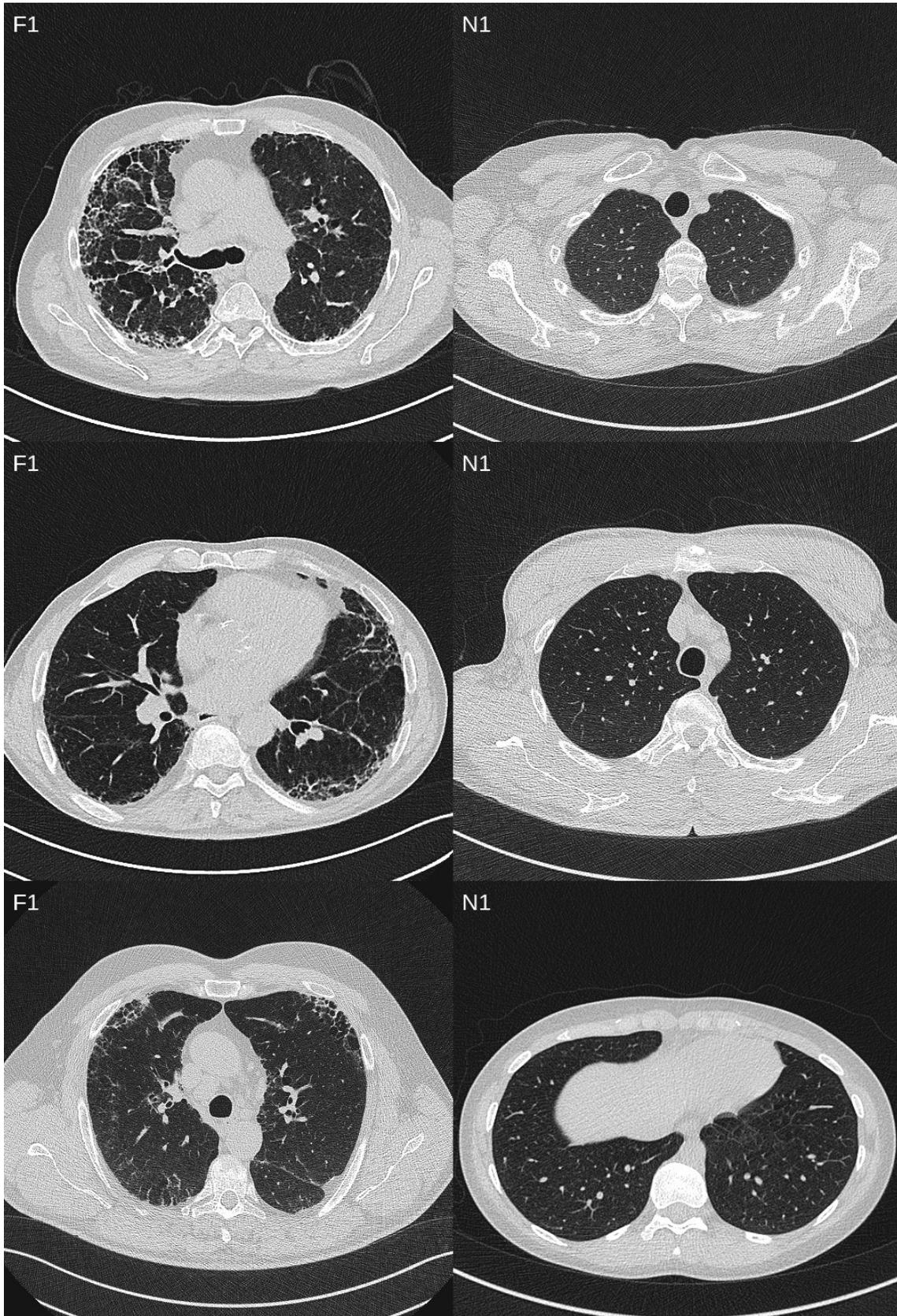


Abbildung 14: richtig klassifiziert, Zuordnungswahrscheinlichkeit 1 (TensorFlow, Inception v3, Variante 2)

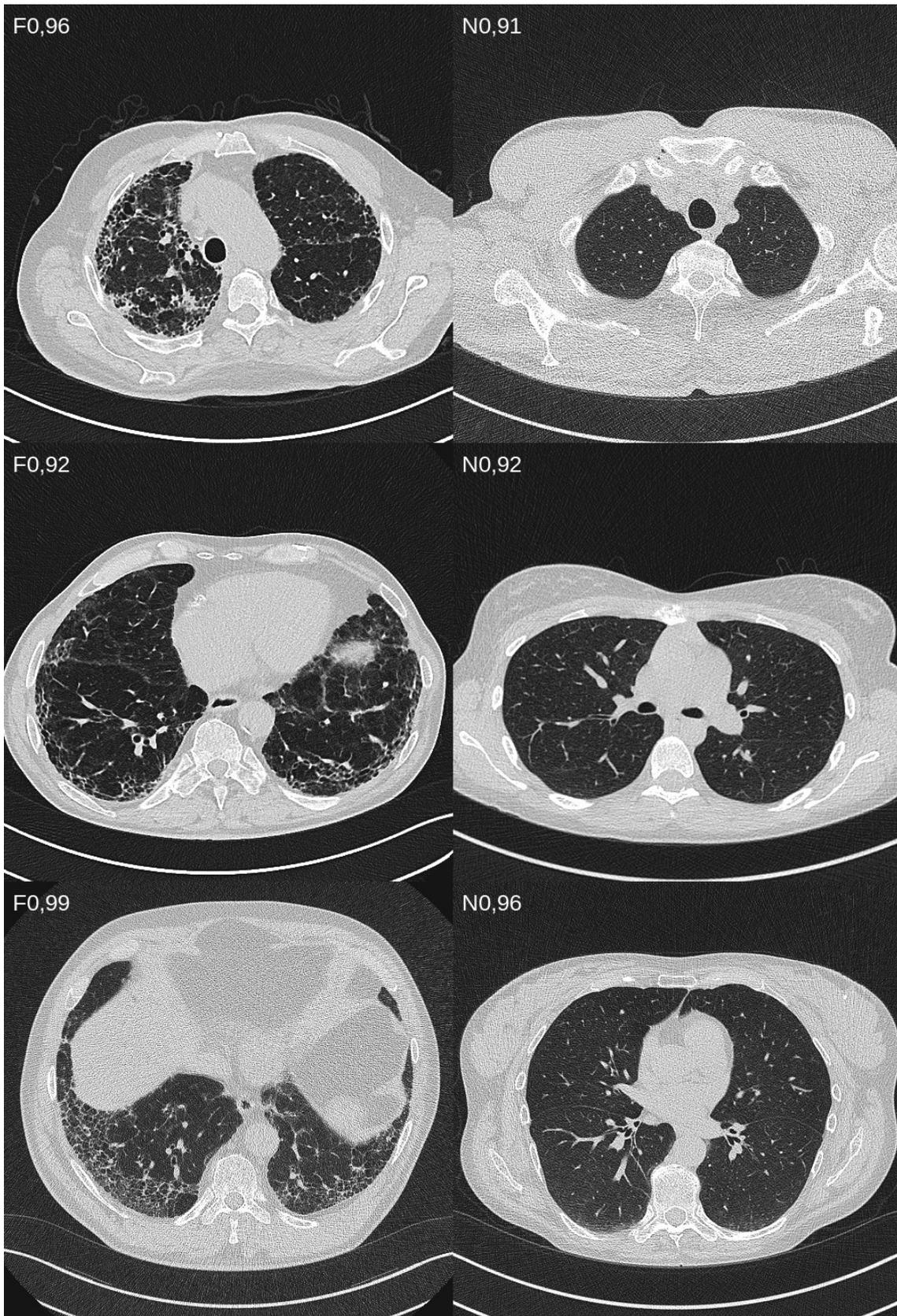


Abbildung 15: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,9 - 0,99 (TensorFlow, Inception v3, Variante 2)

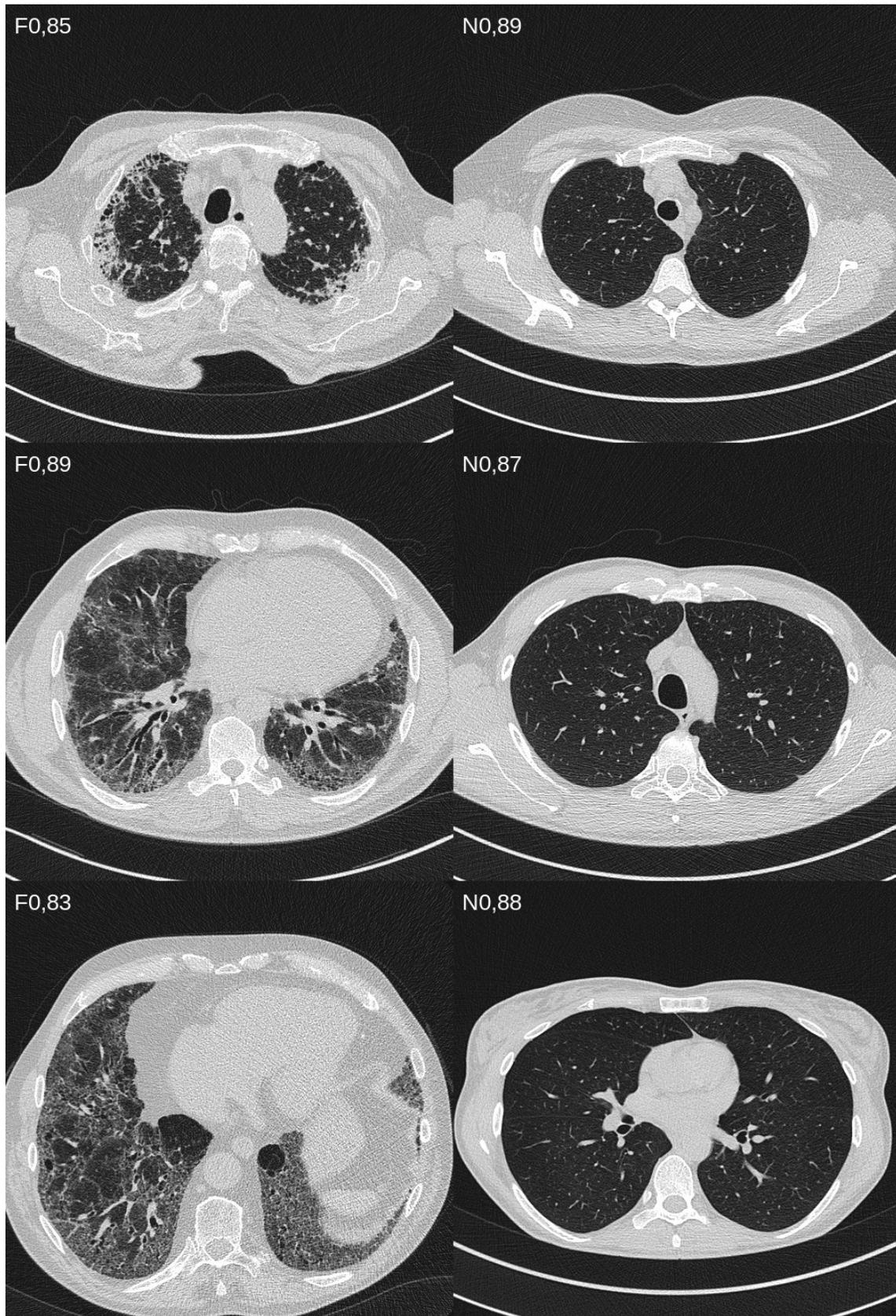


Abbildung 16: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,8 - 0,89 (TensorFlow, Inception v3, Variante 2)

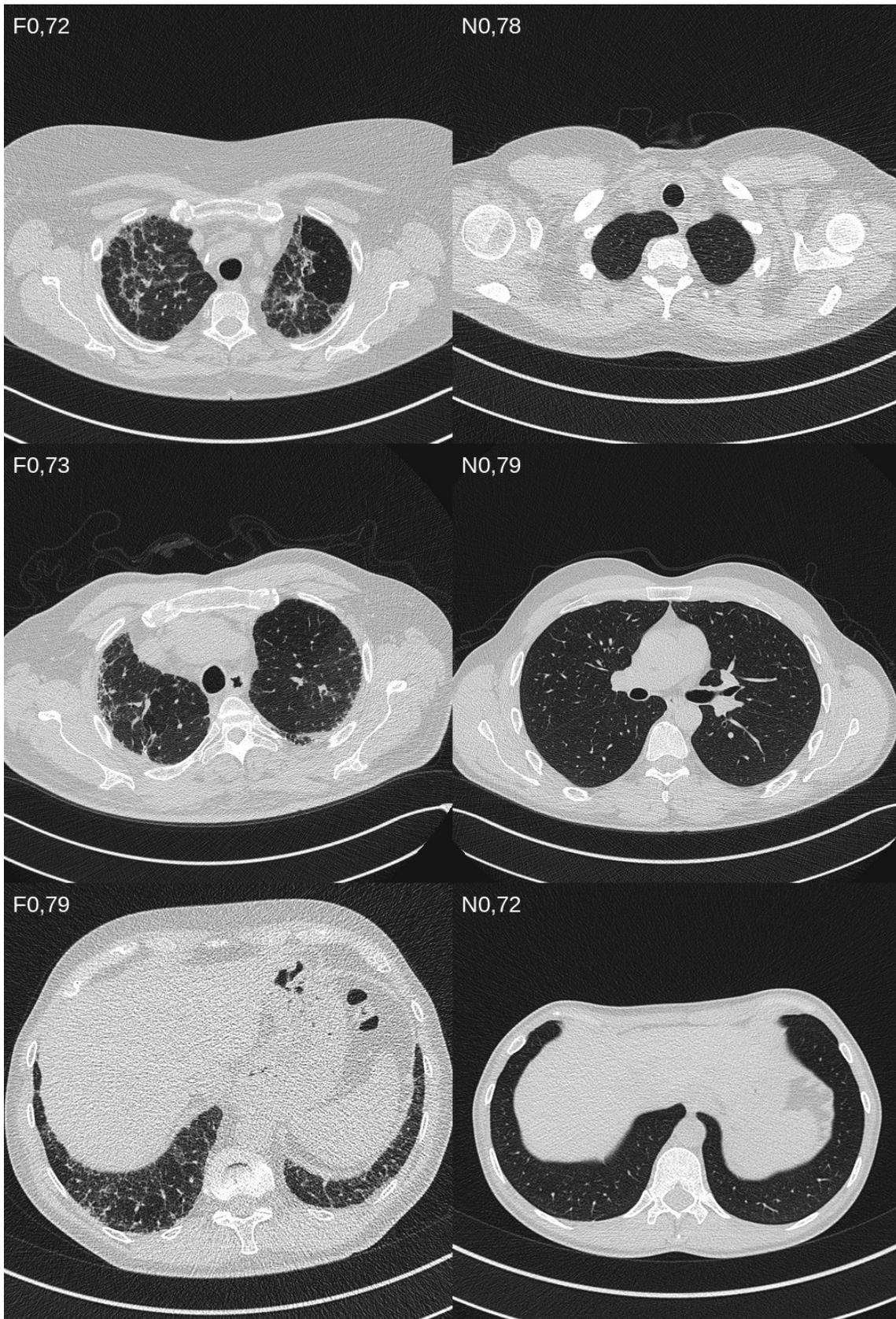


Abbildung 17: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,7 - 0,79 (TensorFlow, Inception v3, Variante 2)

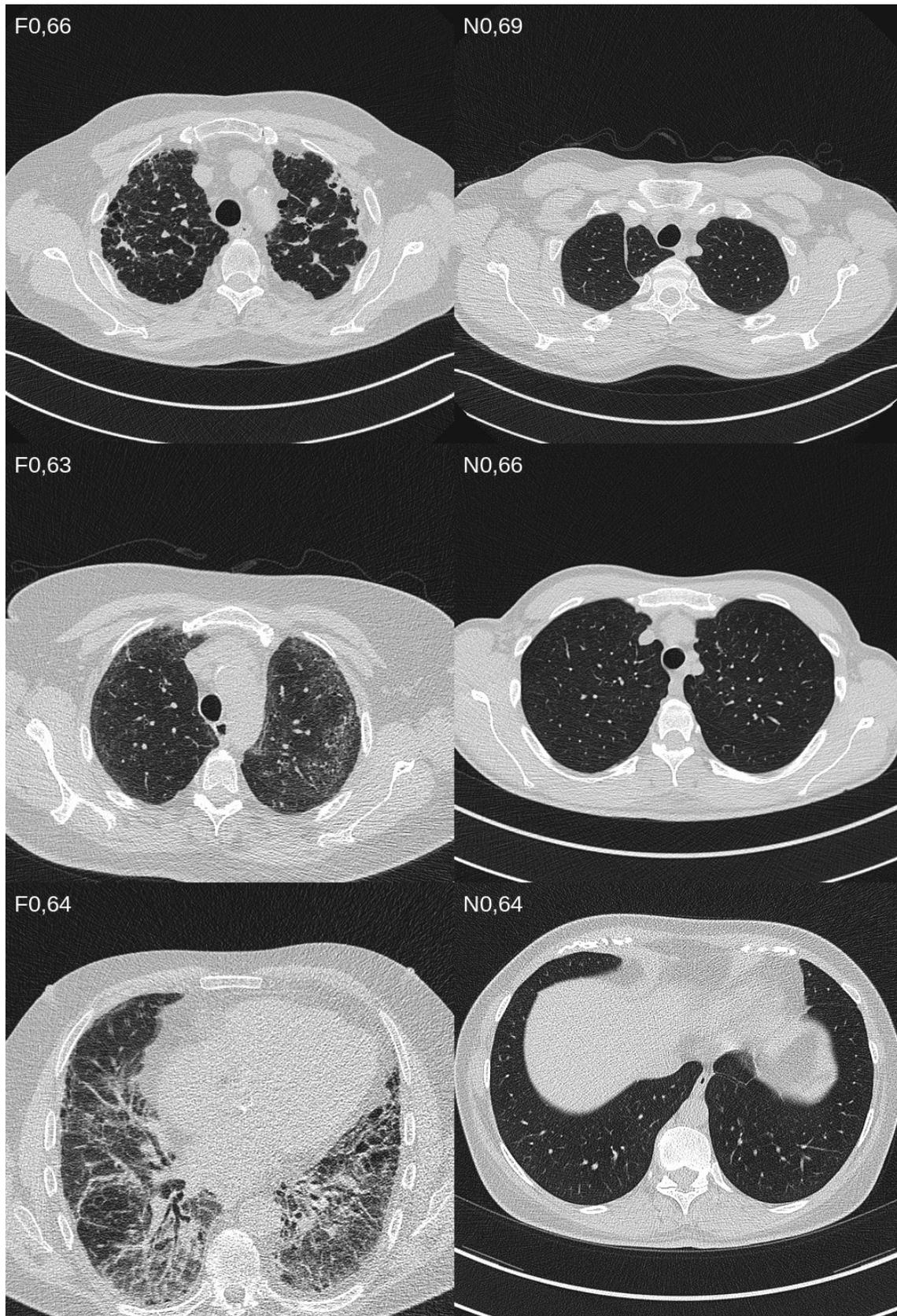


Abbildung 18: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,6 - 0,69 (TensorFlow, Inception v3, Variante 2)

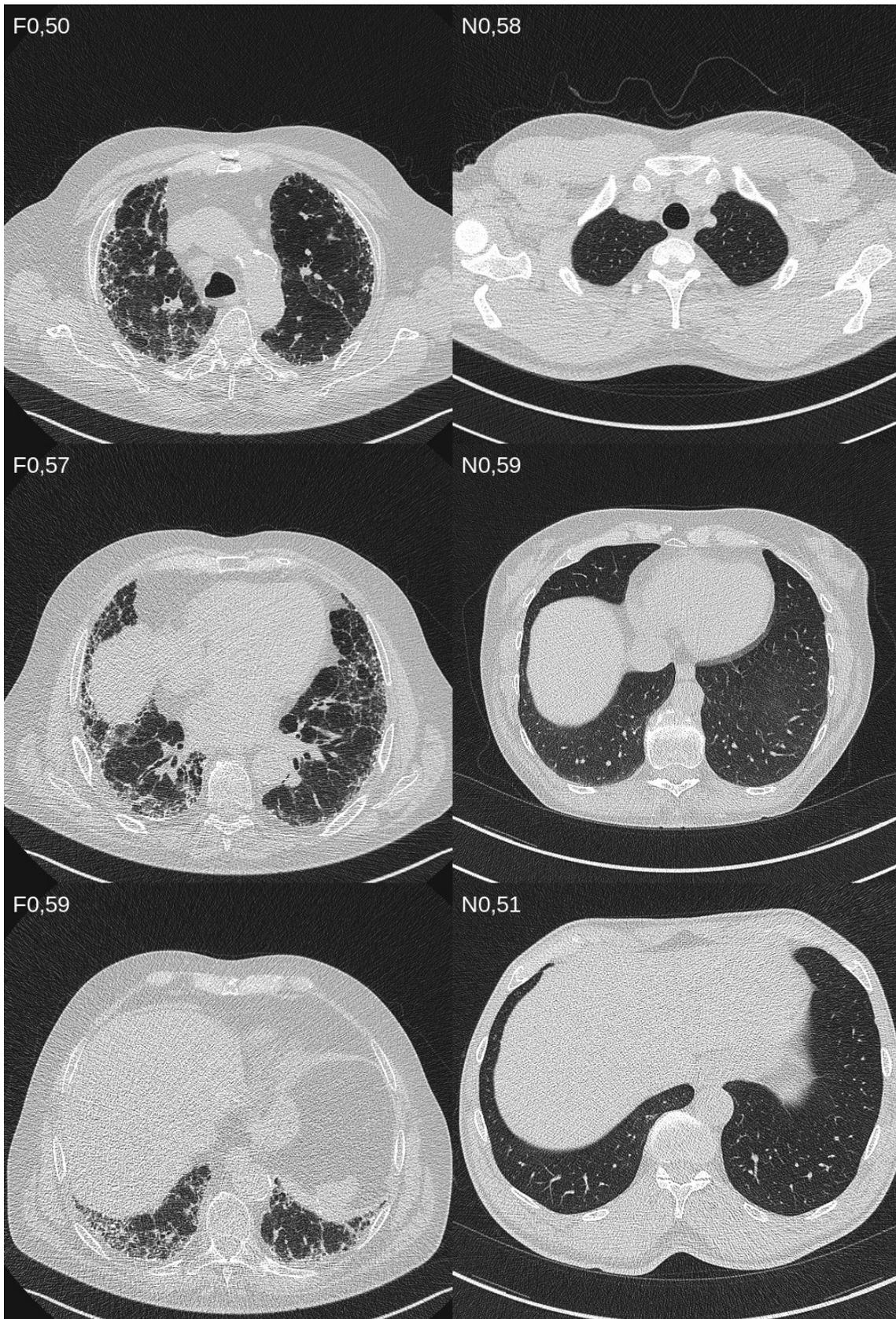


Abbildung 19: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,5 - 0,59 (TensorFlow, Inception v3, Variante 2)

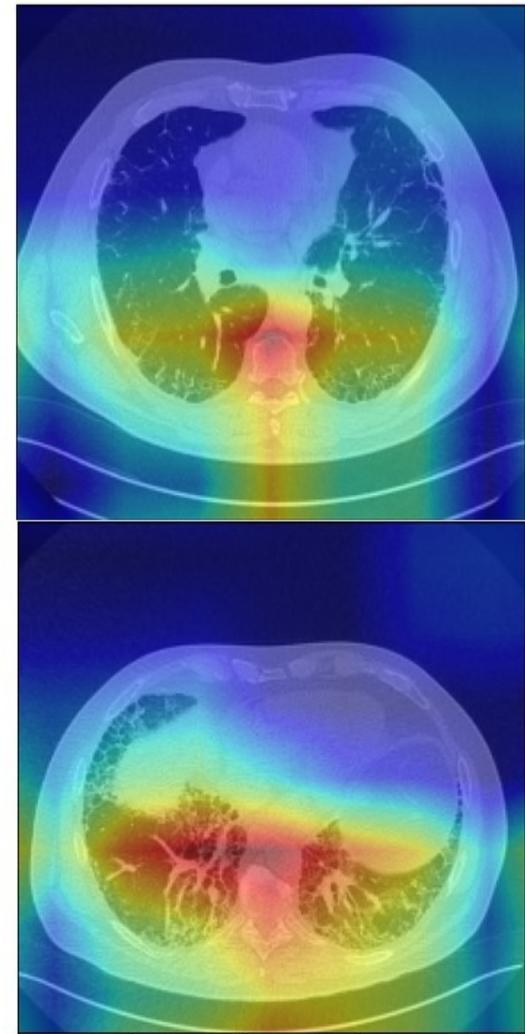
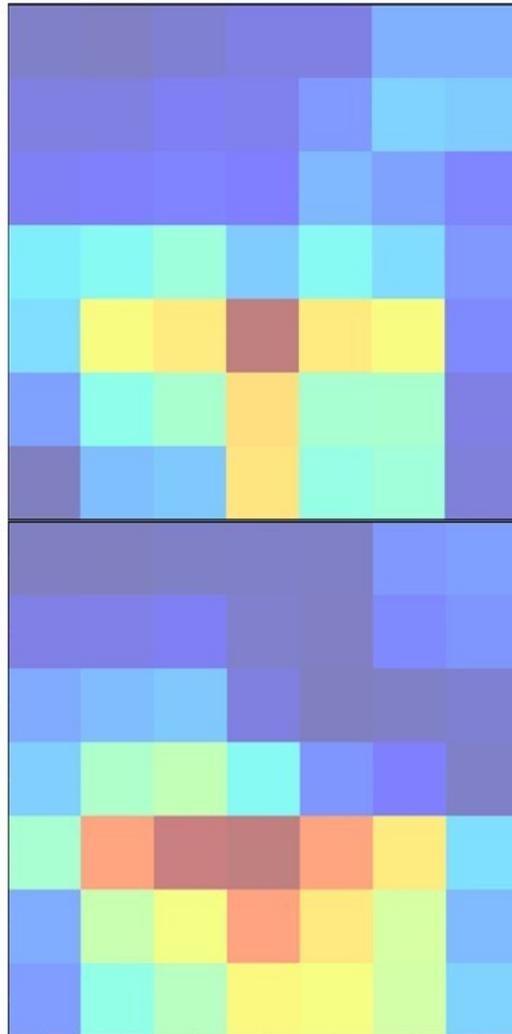
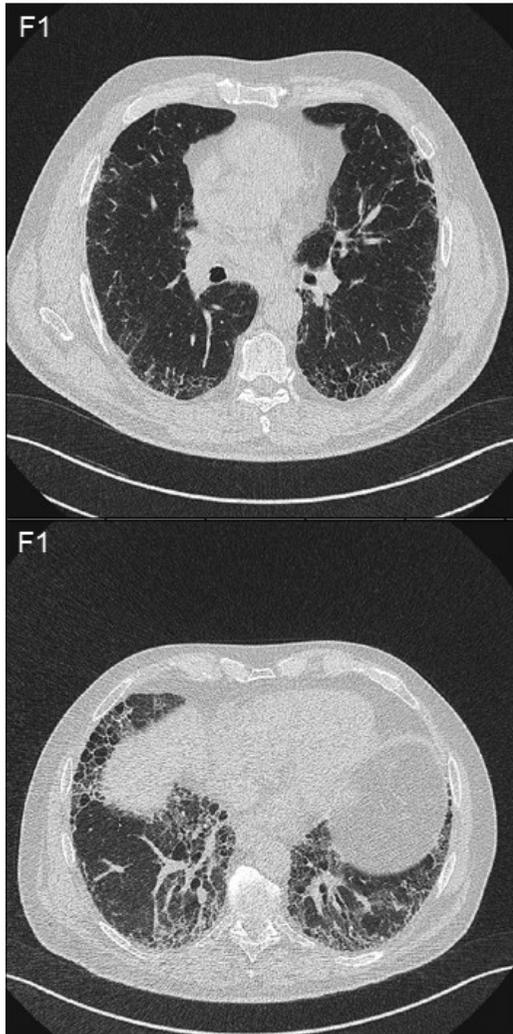


Abbildung 20: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, DenseNet121)

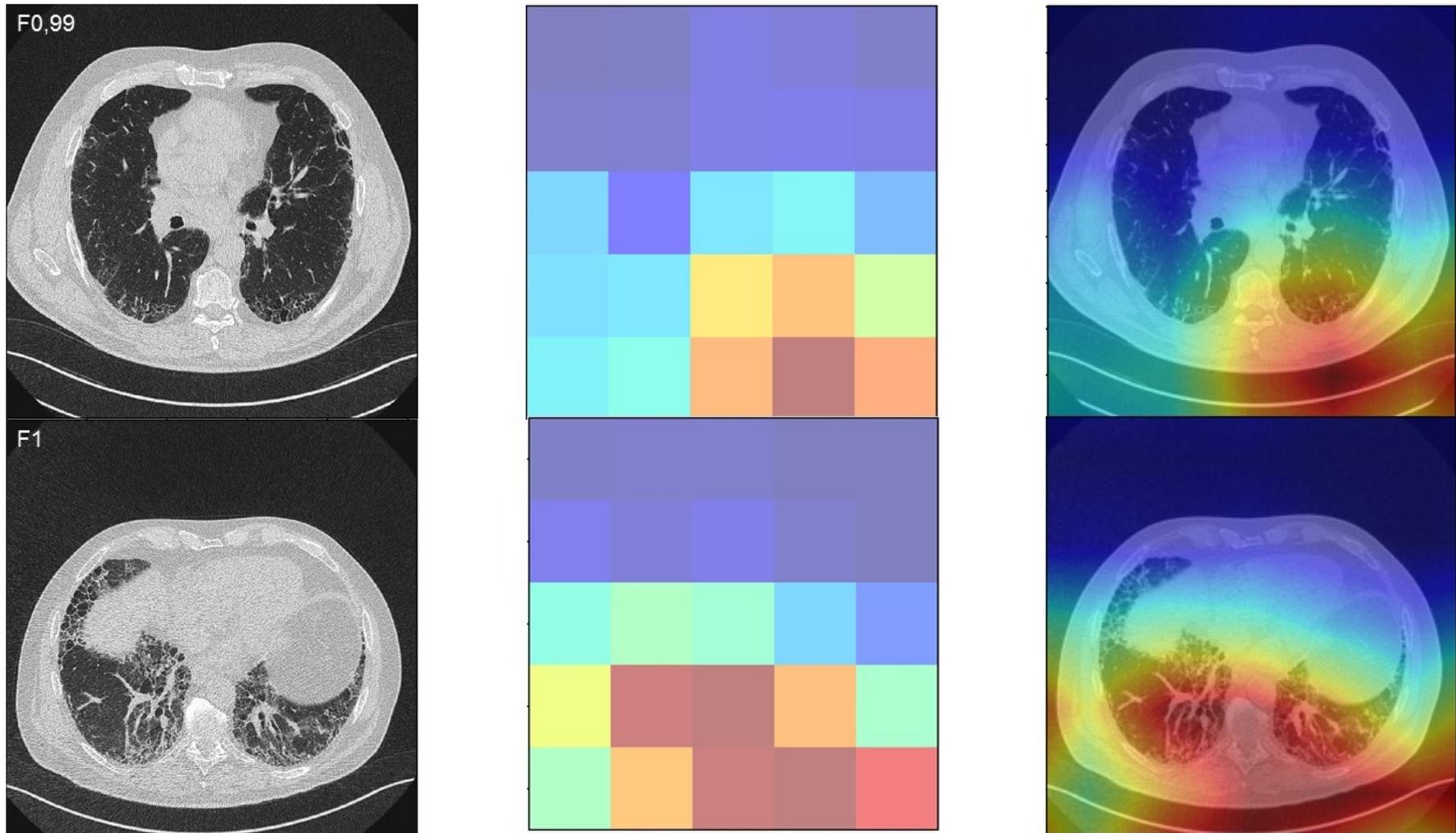


Abbildung 21: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, Inception v3)

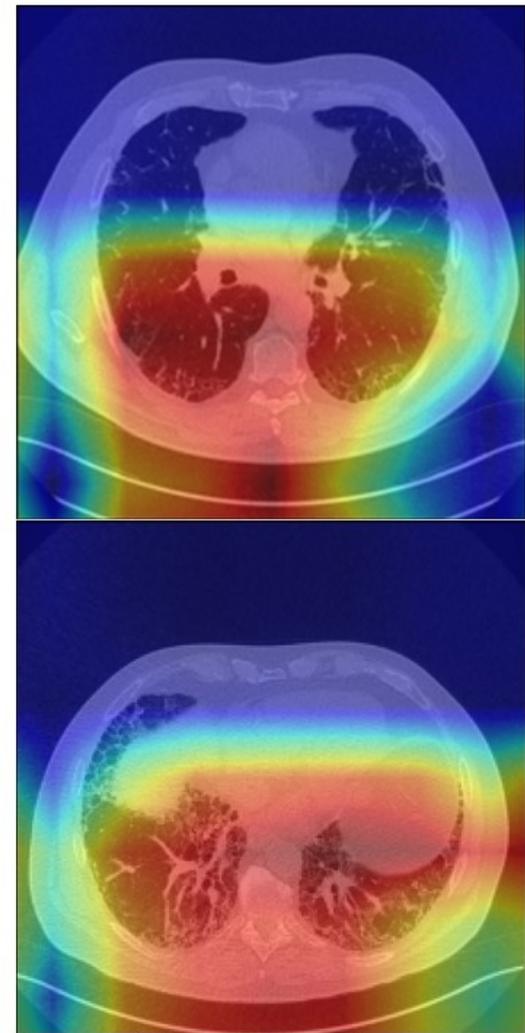
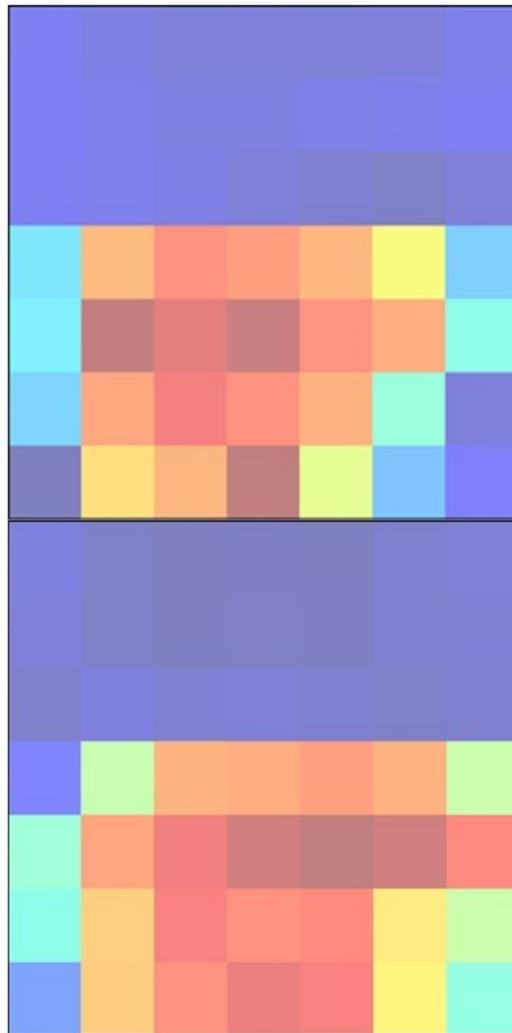
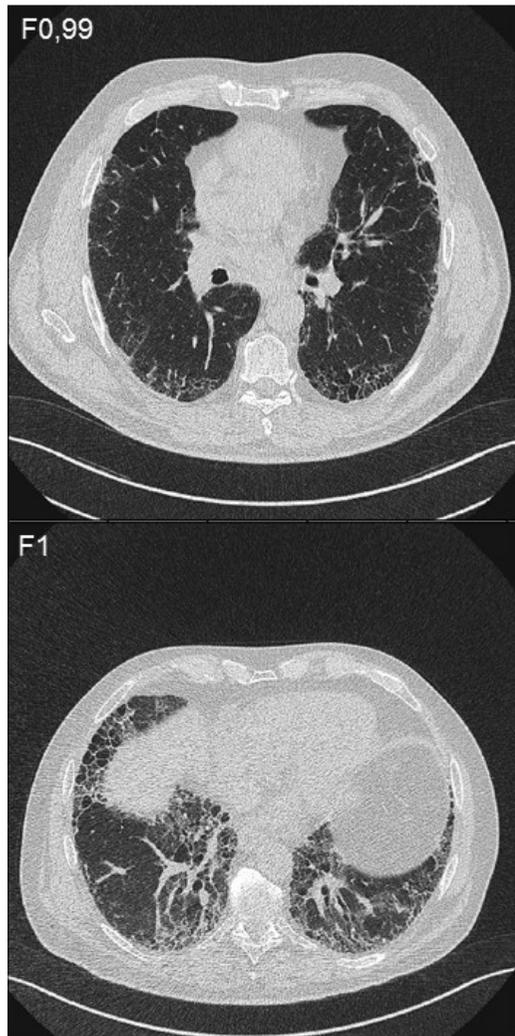


Abbildung 22: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, ResNet50)

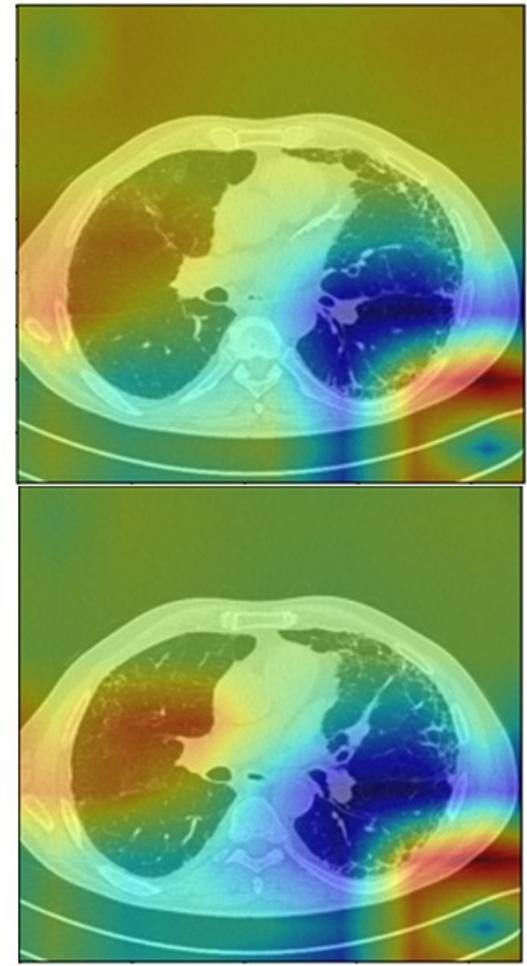
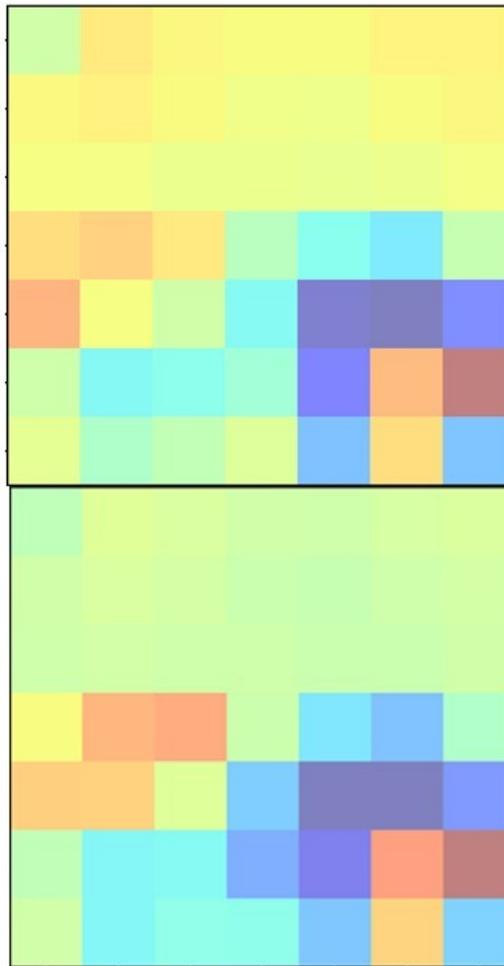
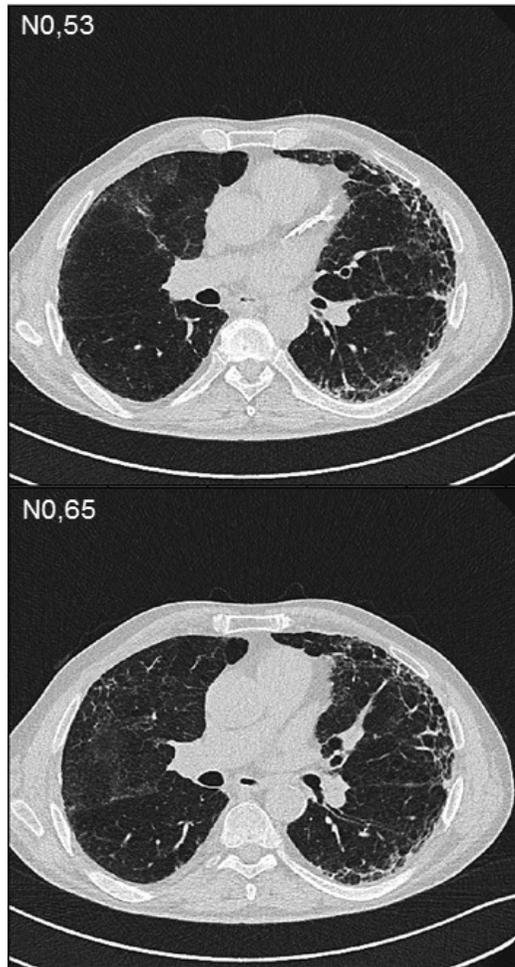


Abbildung 23: falsch klassifiziert mit zugehörigen Heat Maps (PyTorch, ResNet50)

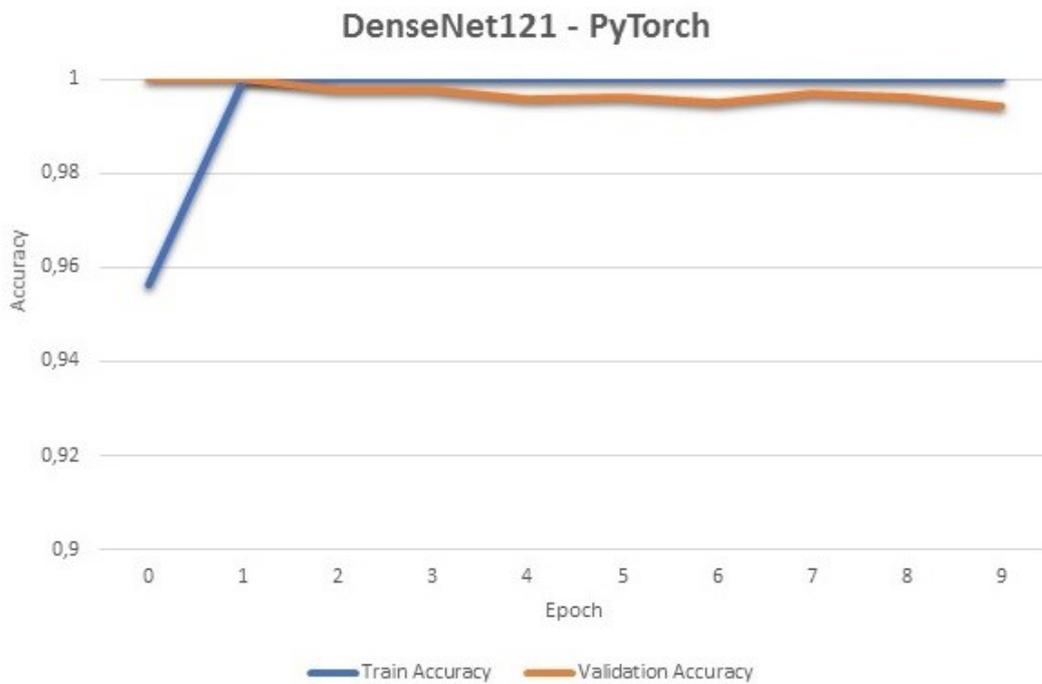


Abbildung 24: Lern- und Validierungsgenauigkeit während der Anlernphase - DenseNet121

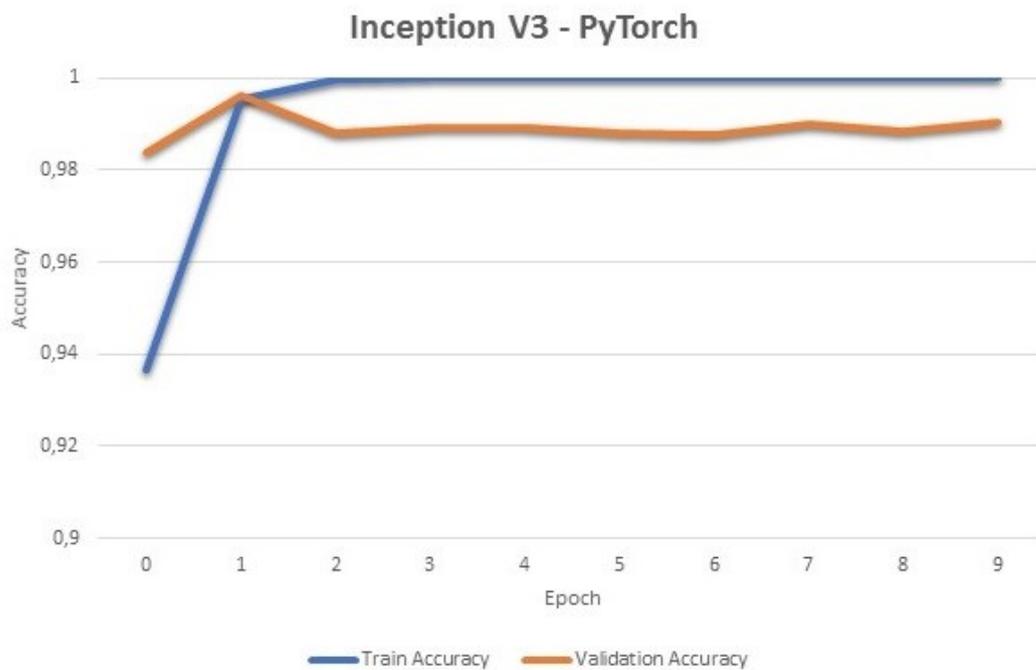


Abbildung 25: Lern- und Validierungsgenauigkeit während der Anlernphase - Inception v3

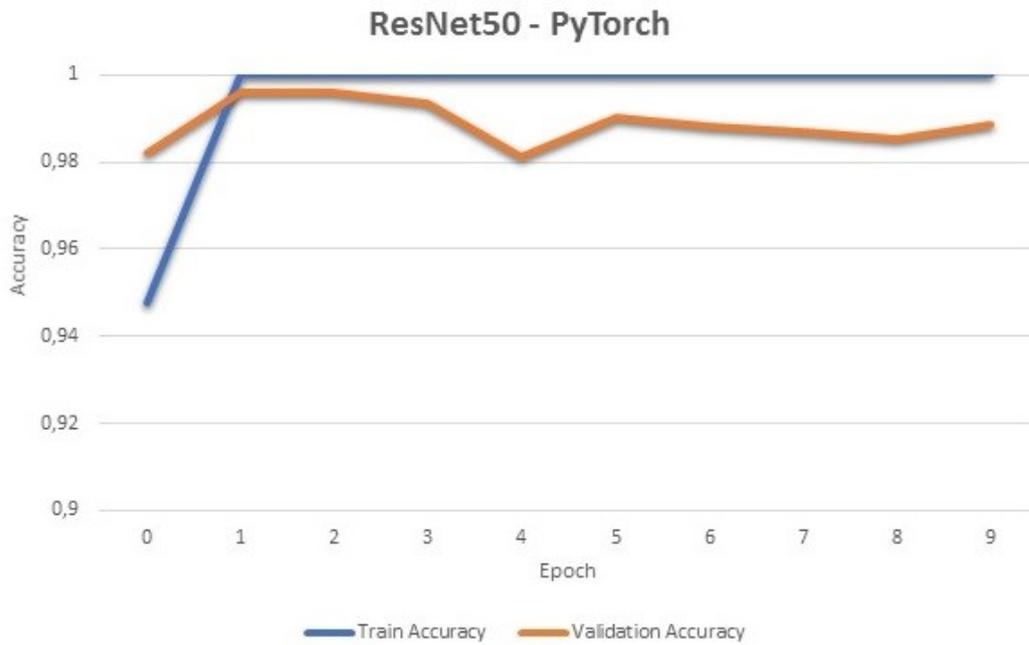


Abbildung 26: Lern- und Validierungsgenauigkeit während der Anlernphase - Res-Net50

Tabelle 1: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (TensorFlow, Inception v3, Variante 1)

Bereich Klassifizierungswahrscheinlichkeit S	Gesund		Fibrose		Spezifität	Sensitivität
	Bildmaterial gesamt	Korrekt zugeordnet	Bildmaterial gesamt	Korrekt zugeordnet		
> 0,99	3	3	69	69	100 %	100 %
0,90 - 0,99	918	914	908	895	99,6 %	98,6 %
0,80 - 0,89	121	113	32	29	93,4 %	90,6 %
0,70 - 0,79	55	44	23	19	80 %	82,6 %
0,60 - 0,69	43	25	14	8	58,1 %	57,1 %
0,50 - 0,59	29	16	11	4	55,2 %	36,4 %
Gesamt Summen	1169	1115	1057	1024	95,4 %	96,9 %

Tabelle 2: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (TensorFlow, Inception v3, Variante 2)

Bereich Klassifizierungswahrscheinlichkeit S	Gesund		Fibrose		Spezifität	Sensitivität
	Bildmaterial gesamt	Korrekt zugeordnet	Bildmaterial gesamt	Korrekt zugeordnet		
> 0,99	4206	4206	2513	2513	100 %	100 %
0,90 - 0,99	856	852	312	312	99,5 %	100 %
0,80 - 0,89	216	209	43	35	96,8 %	81,4 %
0,70 - 0,79	66	61	22	17	92,4 %	77,3 %
0,60 - 0,69	37	29	13	8	78,4 %	61,5 %
0,50 - 0,59	26	16	16	7	61,5 %	43,8 %
Gesamt Summen	5407	5373	2919	2892	99,4 %	99,1 %

Tabelle 3: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, DenseNet121)

Bereich Klassifizierungswahrscheinlichkeit S	Gesund		Fibrose		Spezifität	Sensitivität
	Bildmaterial gesamt	Korrekt zugeordnet	Bildmaterial gesamt	Korrekt zugeordnet		
> 0,99	5394	5394	2811	2811	100 %	100 %
0,90 - 0,99	13	13	92	92	100 %	100 %
0,80 - 0,89	0	0	13	13	NA	100 %
0,70 - 0,79	0	0	2	2	NA	100 %
0,60 - 0,69	0	0	1	1	NA	100 %
0,50 - 0,59	0	0	0	0	NA	NA
Gesamt Summen	5407	5407	2919	2919	100 %	100 %

Tabelle 4: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, Inception v3)

Bereich Klassifizierungswahrscheinlichkeit S	Gesund		Fibrose		Spezifität	Sensitivität
	Bildmaterial gesamt	Korrekt zugeordnet	Bildmaterial gesamt	Korrekt zugeordnet		
> 0,99	5115	5115	2664	2664	100 %	100 %
0,90 - 0,99	248	248	223	223	100 %	100 %
0,80 - 0,89	37	37	18	18	100 %	100 %
0,70 - 0,79	7	7	10	10	100 %	100 %
0,60 - 0,69	0	0	4	4	NA	100 %
0,50 - 0,59	0	0	0	0	NA	NA
Gesamt Summen	5407	5407	2919	2918	100 %	100 %

Table 5: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, ResNet50)

Bereich Klassifizierungswahrscheinlichkeit S	Gesund		Fibrose		Spezifität	Sensitivität
	Bildmaterial gesamt	Korrekt zugeordnet	Bildmaterial gesamt	Korrekt zugeordnet		
> 0,99	5371	5371	2233	2233	100 %	100 %
0,90 - 0,99	32	32	441	441	100 %	100 %
0,80 - 0,89	4	4	89	89	100 %	100 %
0,70 – 0,79	0	0	67	67	NA	100 %
0,60 – 0,69	0	0	48	47	NA	97,92 %
0,50 - 0,59	0	0	41	40	NA	97,56 %
Gesamt Summen	5407	5407	2919	2917	100 %	99,93 %

Diskussion

Die verschiedenen neu angelernten Neuronale Netze konnten erfolgreich eingesetzt werden und zeigten beeindruckende Ergebnisse. Die Klassifizierungsergebnisse verbesserten sich bei Erhöhung der Lerndaten und bei patientenspezifischer Zuordnung der Lern- bzw. der Testdaten bei gleichbleibender Programmbibliothek (TensorFlow). Im Vergleich dazu zeigte das neu angelernte Neuronale Netz Inception v3 unter PyTorch (Programmbibliothek) bessere Klassifizierungsergebnisse als unter TensorFlow. Als Hauptgrund hierfür ist die unterschiedliche Datenstruktur zu nennen: Unter PyTorch wurden die Daten patientenspezifisch in drei Datensätze aufgeteilt: Lerndaten, Validierungsdaten und Prüfdaten. Dabei wurde sichergestellt, dass sich sowohl Patienten, als auch die Schnittbilder in den einzelnen Datengruppen nicht überschneiden. In der TensorFlow-Umgebung hingegen erfolgte eine patientenspezifische Aufteilung der Daten lediglich in zwei Datensätze: Lehrdaten (Lern- und Validierungsdaten) und Prüfdaten. Hierbei wurde nicht sichergestellt, dass sich Patienten in Lern- und Validierungsdaten nicht überschneiden. Eine patienten- und schnittbilderspezifische Aufteilung der Daten erscheint deshalb sinnvoll.

Die Fehlzusammenordnungen (Abbildung 12 und 13) in TensorFlow sind praktisch nicht nachvollziehbar („Black Box“). Somit erscheinen die Klassifizierungsergebnisse nicht transparent. „Explainable AI“ beschäftigt sich mit Methoden, um die Entscheidungsfindung Neuronaler Netze besser nachvollziehen zu können (Abbildung 27). Einer dieser Methoden ist die Erstellung von Class-Activation-Maps (CAM, Heat Maps). Die Abbildungen 20-23 zeigen die dazugehörigen Heat Maps und ermöglichen einen Einblick in die Entscheidungsfindung der Neuronalen Netze in PyTorch. ResNet50 klassifizierte zwei fibrotische Bilder als „Gesund“ (Abbildung 23). Beide Schnittbilder sind vom selben Patienten und zeigen fibrotische Lungengerüstveränderungen der linken Lunge. Die rechte Lunge ist nicht betroffen. Die dazugehörigen Heat Maps zeigen, dass die Entscheidungsfindung „Gesund“ sich vor allem auf die rechte Lunge bezieht. DenseNet121 und Inception v3 zeigten unter PyTorch eine Sensitivität und eine Spezifität von 100 %, wobei durch DenseNet121 mehr Daten mit einer höheren Zuordnungswahrscheinlichkeit klassifiziert wurden (Tabelle 5). Auch in der Lernphase ist DenseNet121 den anderen Neuronalen Netzen in Lern- und Validierungsgenauigkeit überlegen (Abbildung 24). Es ist eine verlässliche Generalisierbarkeit der angelernten Modelle, insbesondere von DenseNet121 in Hinblick auf die Fragestellung nach fibrotischen Lungengerüstveränderungen anzunehmen. Bei jedoch fehlendem multizentrischen Prüfdatensatz kann die Generalisierbarkeit nur eingeschränkt bestätigt werden.

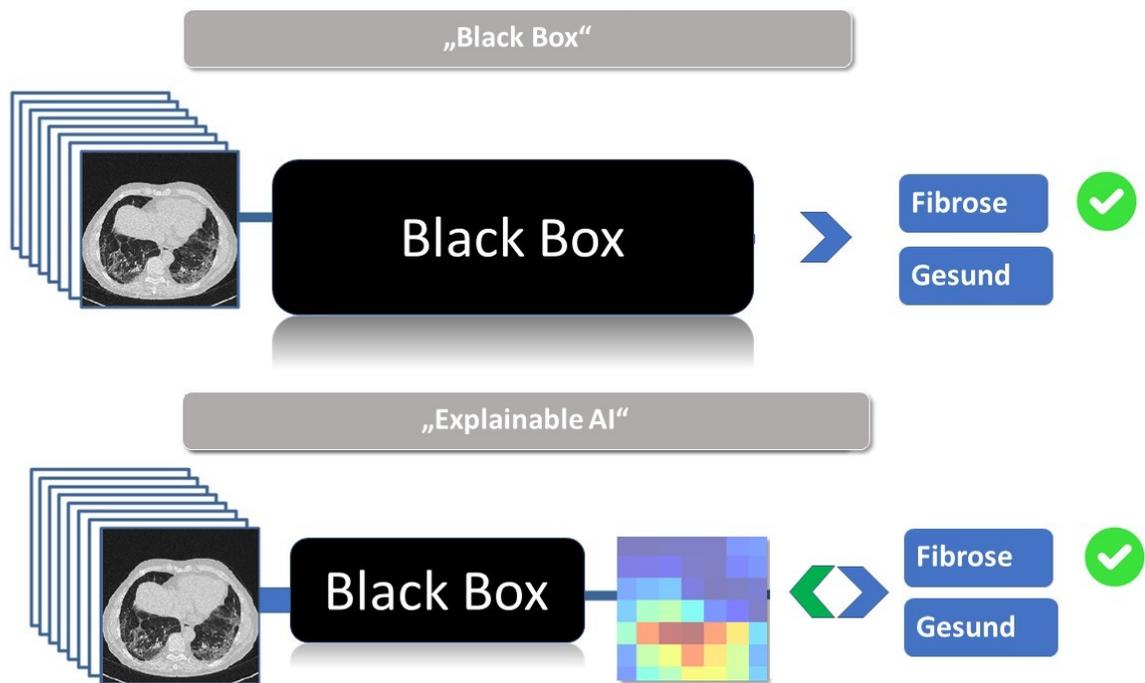


Abbildung 27: Explainable AI

Welchen Vorteil bietet maschinelles Lernen gegenüber regelbasierten Algorithmen?

Um regelbasierte Algorithmen für die Klassifizierung von CT-Schnittbildern in Gesund und Fibrose aufzustellen bedarf es mehrerer komplexer Programmierschritte, unter anderem gegebenenfalls einer Lungensegmentierung. Eine genaue Lungensegmentierung, gerade bei Patienten mit interstitiellen Lungenerkrankungen, wie der Lungenfibrose, ist eine herausfordernde und schwierige Aufgabe. Als nächstes müssen „Gesund“ und „Fibrose“ mathematisch und anhand von Regeln definiert werden. Die Klassifizierung ist keine triviale Aufgabe. Die Abbildung 28 stellt einen vereinfachten Ablauf regelbasierter Methoden dar.

Im Gegensatz zu regelbasierten Algorithmen lernt die künstliche Intelligenz (KI) aus Lerndaten. Es müssen keine festen Regeln definiert werden. Die KI sucht nach Gemeinsamkeiten und Unterschieden in den Lerndaten und definiert somit selbst Muster (Patterns). Abbildung 29 stellt einen vereinfachten Ablauf maschinellen Lernens dar.

Sehr komplexe Bereiche, für die bisher keine regelbasierten Algorithmen existieren, können mittels maschinellen Lernens analysiert werden, wie beispielsweise Sprach- und Bilderkennung (Mikolov, Deoras et al. 2011, Hinton, Srivastava et al. 2012, Sainath, Kingsbury et al. 2015).

Zudem kann maschinelles Lernen Menschen dabei helfen, Daten bzw. ein Ergebnis besser zu verstehen (Abbildung 30).

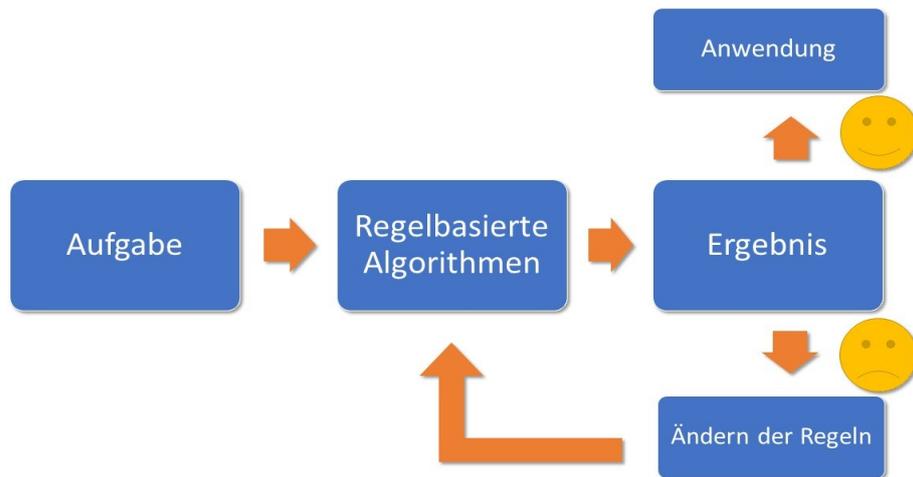


Abbildung 28: regelbasierte Methoden-Eine Aufgabe wird mittels mathematisch definierter Regeln gelöst. Ist das Ergebnis zufriedenstellend, kann das Modell in Anwendung gebracht werden. Sollte das Ergebnis nicht zufriedenstellend sein, sollten die zuvor mathematisch definierten Regeln überdacht bzw. geändert werden.

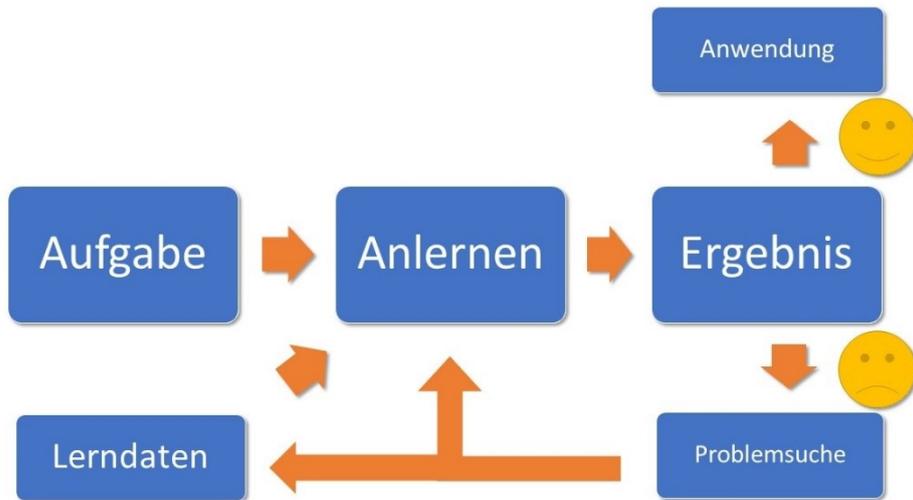


Abbildung 29: maschinelles Lernen-Ein Modell wird anhand Anlerndaten und einem bestimmten Anlernalgorithmus im Hinblick auf eine bestimmte Fragestellung angelernt. Ist das Ergebnis des Modells zufriedenstellend, kann es in Anwendung gebracht werden. Ist das Ergebnis nicht zufriedenstellend, sollten die Anlerndaten auf eine ausreichende Quantität/Qualität bzw. der Anlernalgorithmus auf Fehler überprüft werden.

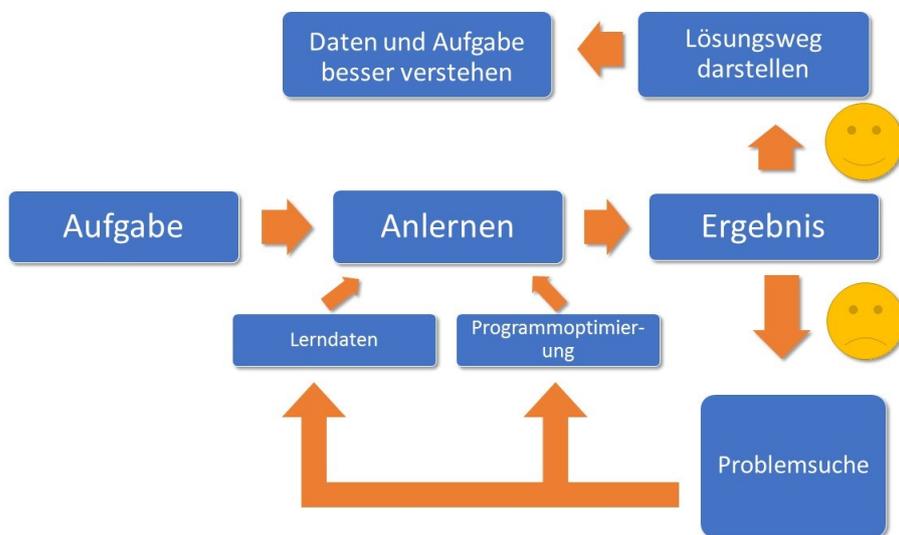


Abbildung 30: Daten besser verstehen-Ein Modell wird anhand Anlerndaten und einem bestimmten Anlernalgorithmus im Hinblick auf eine bestimmte Fragestellung angelernt. Ist das Ergebnis des Modells zufriedenstellend, kann der Lösungsweg dargestellt werden. Dies kann zum besseren Verständnis der Daten beitragen. Ist das Ergebnis nicht zufriedenstellend, sollten die Anlerndaten bzw. der Anlernalgorithmus auf suboptimale Lösungsansätze überprüft werden.

Limitationen dieser Studie

Die Gruppenanzahl, in die klassifiziert wird, ist bei dieser Untersuchung sehr klein (binär) und besteht nur aus Gesund und Fibrose. Radiologen hingegen klassifizieren zwischen hunderten von verschiedenen Erkrankungen, Mischformen verschiedener Pathologien und dies inklusive unterschiedlicher Schweregrade. Das neu angelernete Neuronale Netz kann nur zwischen Gruppen differenzieren, für die es angelernt wurde. Es besteht keine kontextübergreifende Intelligenz. Dem Radiologen kommt bei der Entwicklung von KI Software eine besondere Bedeutung zu, denn die Bereitstellung von guten Lern- und Prüfdaten erfordert Expertenwissen und ist, insbesondere durch Mischdiagnosen, sehr komplex und zeitaufwendig. Es ist überhaupt nicht trivial das radiologische Detailwissen in eine KI-Software zu implementieren, das die unerschöpfliche Vielfalt an Kombination von verschiedenen Morphologien beinhaltet. Eine Kooperation zwischen vielen Radiologen und IT Experten scheint daher sinnvoll zu sein.

Die Software erkennt nur wozu sie angelernt wurde. Betrachtet man beispielsweise die Differenzierung unterschiedlicher Ausprägungsgrade einer Fibroseerkrankung, so könnte die KI-Software diese nur erkennen, wenn sie mit Beispieldaten von Gruppen unterschiedlicher Fibrose-Ausprägung zuvor angelernt worden wäre. Die Klassifizierungswahrscheinlichkeit S gibt die Wahrscheinlichkeit für die korrekte Zuordnung zu einer angelernt Gruppe an. Sie korreliert nicht mit dem Ausprägungsgrad einer Bildeigenschaft.

Deep Learning Algorithmen stellen lediglich eine objektive Beurteilung eines Bilddatensatzes dar, ohne jegliche Korrelation mit der Klinik des Patienten. Diese klinisch-pathologische Korrelation ist jedoch für weitere diagnostische Schritte, Interventionen und letztendlich für Therapieentscheidungen unabdingbar.

Diese Algorithmen brauchen viele Anlern Daten. In dieser Arbeit wurde gleich zu Beginn mit Anlern bildmengen von ca. 2300 bzw. 8950 Bildern begonnen. Dabei wurden direkt hohe korrekte Zuordnungsraten gefunden. In der dermatologischen Arbeit von Esteva und Mitarbeitern wurden insgesamt 129.450 Anlern bilder benutzt, da ein wesentlich komplexeres Klassifizierungsthema aus 2032 verschiedenen Krankheiten bearbeitet wurde (Esteva, Kuprel et al. 2017). Das entspricht 63,7 Bilder pro Erkrankungsklasse.

Inwiefern akzeptable Spezifitäten und Sensitivitäten auch bei Verwendung von wesentlich geringeren Bildanzahlen erreicht werden können, muss in weiteren Arbeiten untersucht werden. Gerade für selteneren Erkrankungen mit wenigen zur Verfügung stehenden Daten, erscheint es sicherlich für sinnvoll klassische Radiomics-Marker mit Deep Learning Algorithmen zu kombinieren (Obert 2019).

Ziel ist es, durch solche Arbeiten KI-Software zu entwickeln, die in Zukunft Bilddaten eigenständig nach verschiedensten Fragestellungen klassifizieren und so momentane ärztliche Tätigkeiten unterstützen kann. Bei der Schweregradeinteilung einzelner Erkrankungen könnte solche Software wahrscheinlich schon bald zur Therapiekontrolle in der Klinik zur Verfügung stehen.

Solange jedoch nicht erklärbare Fehlklassifikationen auftreten wie sie in Abb. 10, 12 oder 22 gezeigt wurden, muss man die 100 % an korrekten Zuordnungen mit großer Vorsicht interpretieren, da 100% richtige Klassifikationen bei den geringen Fallzahlen nicht vertrauenswürdig sind. In der Arbeit von Ardila (Lungenkrebsfrüherkennung) wurden nicht 102 Fälle wie in dieser Arbeit, sondern 42.579 Fälle berücksichtigt. Fallzahlen wie bei Lungentumorerkrankungen stehen jedoch zur Klärung der Thematik der Lungenfibrose nicht zur Verfügung. Trotzdem ist es relevant sich auch der Auswertung von selteneren Erkrankungen mit KI-Methoden zu widmen, um deren Einsatzfähigkeit auch für diese seltener auftretenden Erkrankungen zu evaluieren. Eventuell kann man Methoden des Generative Deep Learning dazu verwenden, um die Fallzahlen zu erhöhen. Als Zweitmeinung sind die gefundenen Resultate für den erfahrenen Radiologen bereits jetzt interessant, wenn man die dazugehörigen „Heat Maps“ mit aufzeigt. Eventuell kann man die Fallzahlen für die KI-Programme bei nicht systemischen und nicht paarig auftretenden Erkrankungen auch noch dadurch erhöhen, dass man die rechte und linke Lunge separat anlernt bzw. auswertet. Hierdurch würde sich sowohl die Anzahl der Lern- bzw. Prüfdaten erhöhen, als auch die Klassifizierungsgenauigkeit. Somit könnten nämlich Pathologien genauer lokalisiert (rechts bzw. links) werden.

Eine weitere Einschränkung der angelernten Modelle ist die Generalisierbarkeit. Die Modelle wurden nicht an einem multizentrischen Datensatz evaluiert. Somit kann keine valide Aussage über eine Generalisierbarkeit getroffen werden. Zudem wurden artefaktreiche CT-Aufnahmen, weder in der Lernphase, noch in der Prüfphase mit einbezogen. Die Frage, wie mit artefaktbehafteten Bildern umzugehen ist, müsste in weiteren Studien untersucht werden.

Strukturierte Befundung

Die tägliche Arbeit eines Radiologen bei der Befundung ist hauptsächlich von zwei Aktivitäten geprägt: die Interpretation einer bildgebenden Studie und die Kommunikation der Ergebnisse an den überweisenden Arzt. In den letzten Jahrzehnten hat es auf dem

Gebiet der bildgebenden Diagnostik erhebliche Fortschritte gegeben. Wobei in der Vergangenheit das konventionelle Röntgen eine Hauptrolle einnahm, kommen in der heutigen modernen Zeit immer mehr Computertomographie- (CT) und Magnetresonanztomographieuntersuchungen (MRT) hinzu. Um die Produktivität zu steigern, wurden über die letzten Jahre viele Arbeitsabläufe vollständig digitalisiert und es wurden elektronische Bildarchivierungs- und Kommunikationssysteme (PACS) sowie die Spracherkennung eingeführt.

Die heutigen radiologischen Berichte sind im Allgemeinen in Unterabschnitte wie „Klinische Informationen“, „Befunde“ und „Beurteilung“ unterteilt, folgen jedoch nicht einer einheitlichen Struktur. Gerade bei all den neuen und aufkommenden Technologien in der künstlichen Intelligenz ist es für die weitere Entwicklung entscheidend, dass radiologische Berichte einer einheitlichen Struktur folgen, damit spezifische Informationen einfacher extrahiert werden können. Leider ist es bis heute relativ schwierig, Informationen in Freitexten zu extrahieren. Es existieren bereits Verfahren, wie zum Beispiel die „Natural Language Processing“ (NLP)-Techniken, die in den letzten Jahren sicherlich erhebliche Verbesserungen gezeigt haben, jedoch in Freitexten nicht ausreichend überzeugen konnten. Es gibt einige relevante Herausforderungen, die bei der Arbeit mit konventionellen, unstrukturierten Berichten existieren: Unterschiede in Sprache, Stil und Befunde, die manchmal sehr unterschiedlich formuliert werden. Berichte von selbst einem Radiologen zu gleichen klinischen Fragen können erheblich variieren, was zu Inkonsistenzen führt. Es erscheint daher sinnvoll, auf standardisierte und strukturierte Berichte umzusteigen. Diese sind objektiver und ermöglichen eine gezielte Extraktion von Informationen, welche wiederum von externen, zum Beispiel KI-basierten Anwendungen genutzt werden können.

Extrahieren der Informationen mittels Natural Language Processing

Die strukturierte Befundung ist bis heute nicht weitreichend umgesetzt. Darüber hinaus existiert eine massive Anzahl an Freitext-Berichten. Es wäre somit wünschenswert, die in diesen Berichten enthaltenen Informationen retrospektiv zu extrahieren. Nicht nur für retrospektive klinische Studien, sondern auch und gerade für die Entwicklung von KI-Systemen, die stark von vorklassifizierten Daten abhängig sind, wären diese Informationen wertvoll. Natural Language Processing (NLP) zielt auf diese Herausforderung ab und beschreibt Techniken zur Extraktion relevanter Informationen, unter anderem aus Freitexten. Da die meisten Freitexte nur eine begrenzte Struktur aufweisen und je nach Arzt in der Wortwahl variieren, konnten diese Techniken nur bis

zu einem gewissen Grad erfolgreich eingesetzt werden. Nichtsdestotrotz haben die NLP-Technologien in den letzten Jahren erhebliche Verbesserungen gezeigt, vor allem mit dem Einsatz des IBM Watson in der Jeopardy-Sendung (Tesauro, Gondek et al. 2013). Daraufhin wurden zahlreiche Versuche unternommen, Informationen aus klinischen Texten mit verschiedenen Techniken zu extrahieren (Kreimeyer, Foster et al. 2017). Während sich die meisten Studien auf andere klinische Texte als den radiologischen Befund konzentrierten (Ni, Wright et al. 2015), gab es auch einige NLP-Systeme, die speziell entwickelt wurden, um Informationen aus radiologischen Berichten zu extrahieren (Cai, Giannopoulos et al. 2016). Aufgrund der hohen Variabilität des verwendeten Vokabulars war die Leistung traditioneller, regelbasierter Ansätze für NLP nicht optimal, insbesondere wenn Fälle außerhalb des Instituts verarbeitet wurden (Kreimeyer, Foster et al. 2017). Mit der Anwendung neuer Techniken aus dem Gebiet des maschinellen Lernens, zeigten sich deutliche Verbesserungen bei der Extraktion einer Vielzahl von Informationen aus radiologischen Berichten (Gerstmair, Daumke et al. 2012, Hassanpour and Langlotz 2016, Zech, Pain et al. 2018).

Es existieren unglaublich große Datenmengen an radiologischen Berichten, die in Systemen vieler Institutionen verfügbar sind. In der Zukunft werden diese NLP-Technologien sicherlich eine entscheidende Rolle bei der Extraktion wertvoller Informationen spielen, die wiederverwendet werden können, z.B. zum Anlernen neuer KI-Modelle. Allerdings müssen diese vorklassifizierte Daten mit einer gewissen Vorsicht betrachtet werden, wie kürzlich der Fall von CheXNet zeigte (Rajpurkar, Irvin et al. 2017). Hierbei wurde ein Datensatz von etwa 108.000 Röntgenthoraxbildern (Wang, Peng et al. 2017) verwendet, um ein neuronales Netz zur Erkennung von Pneumonien anzulernen. Das NLP-System wurde zur Extraktion von Informationen aus den radiologischen Befunden verwendet. Diese extrahierten Informationen dienten für das neuronale Netz als Klassifizierungsgrundwahrheiten. Aufgrund der enormen Variabilität in den Begriffen, die zur Beschreibung eines bestimmten Befundes verwendet wurden, wiesen die extrahierten Klassifizierungen jedoch einige relevante Unstimmigkeiten auf (Oakden-Rayner 2018). Gerade zum Anlernen neuronaler Netze sind validierte und qualitative Daten essentiell, da das KI-System ansonsten falsche Informationen lernt.

Generelle Aspekte von KI-Methoden

Sobald die KI-Befundungsqualität den Radiologen übertrifft wäre es im Prinzip unethisch, wenn man diese Software nicht zum Wohl des Patienten einsetzen würde.

Selbst wenn man an diesem Punkt angelangt ist, gibt es viele noch zu lösende Herausforderungen wie die Validierung der KI-Software oder auch den Schutz vor böswillig manipulierter Software. Es gibt also noch sehr viele offene Fragen, die auch von Radiologen bearbeitet werden können, die nicht alle IT-Experten sein müssen.

Sollen Patient und Arzt KI-Software vertrauen, die für Einzelne nicht mehr zu durchschauen ist? Wie validiert man komplexe KI-Software? Wer haftet bei Fehlern? Wie schützt man sich vor böswillig manipulierter KI-Software, die systematisch massenhaft Fehldiagnosen verursacht, die eventuell nicht mehr durch Menschen enttarnt werden können? Auch wenn KI-Verfahren beeindruckend sind, so gibt es noch viele Fragen, die vor klinischen Anwendungen grundsätzlich geklärt werden müssen. Es gibt jedoch bereits andere Bereiche, in denen viele dieser Fragen bereits befriedigend gelöst sind. Dies betrifft vornehmlich die Labormedizin mit ihren automatisierten Prozessen.

Angelernte Modelle sollten an einem großen, multizentrischen Testdatensatz getestet werden, ehe sie Anwendung in den klinischen Alltag finden. Werden Modelle an kleinen Datensätzen angelernt besteht das Risiko eines „Overfittings“ (Abbildung 31). Das Modell lernt die Datenpunkte auswendig. Somit ist keine Generalisierbarkeit gegeben. „Underfitting“ ist gegeben, wenn das Model weder die Lerndatenpunkte, noch die Testdatenpunkte erfasst (Abbildung 32). Im besten Fall spiegelt das Modell sowohl die Lern-, als auch die Testdatenpunkte wider (Abbildung 33). Das neuangelernete Neuronale Netz ist in der Lage, das neue Wissen zu generalisieren und kann erfolgreich für die Klassifizierung unbekannter Daten eingesetzt werden.

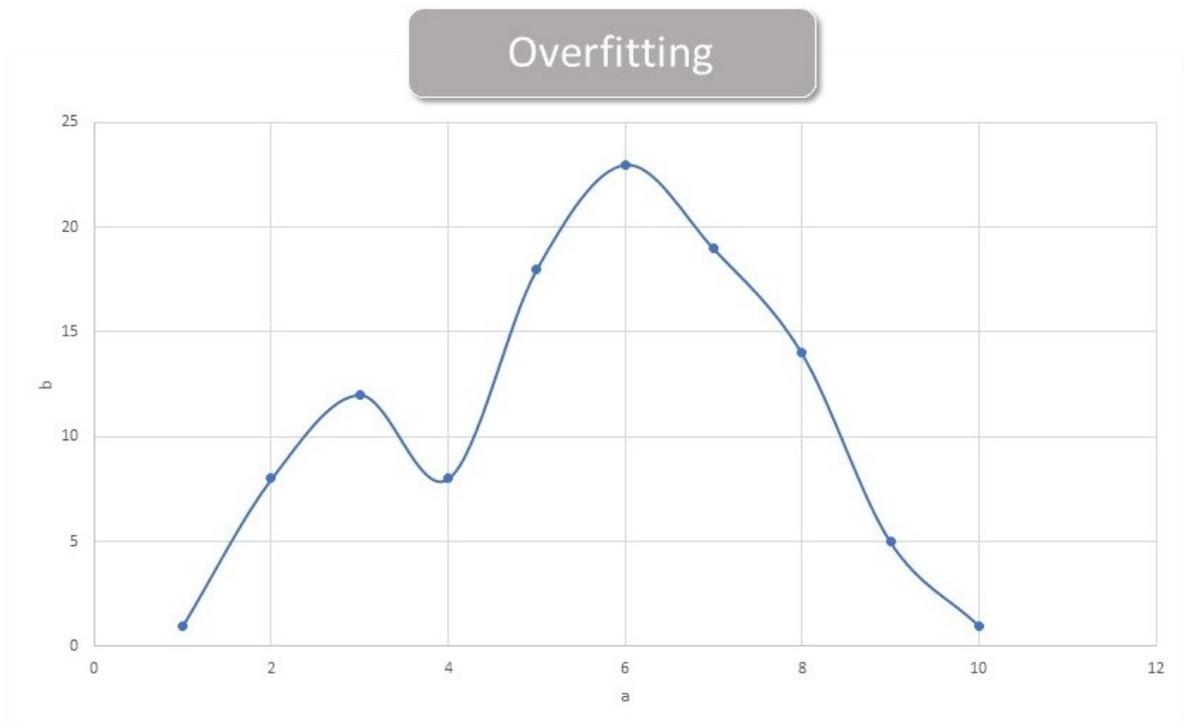


Abbildung 31: Overfitting-gefittete Linie trifft exakt alle Punkte

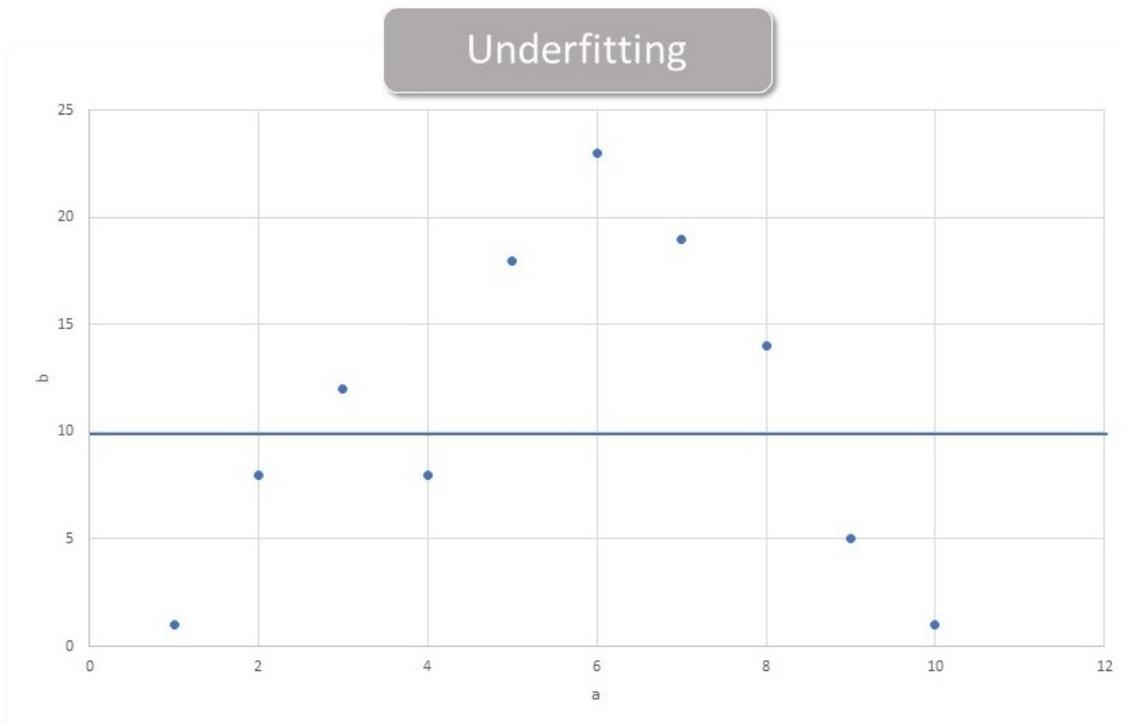


Abbildung 32: Underfitting-gefittete Linie korreliert weder mit Parameter A oder B

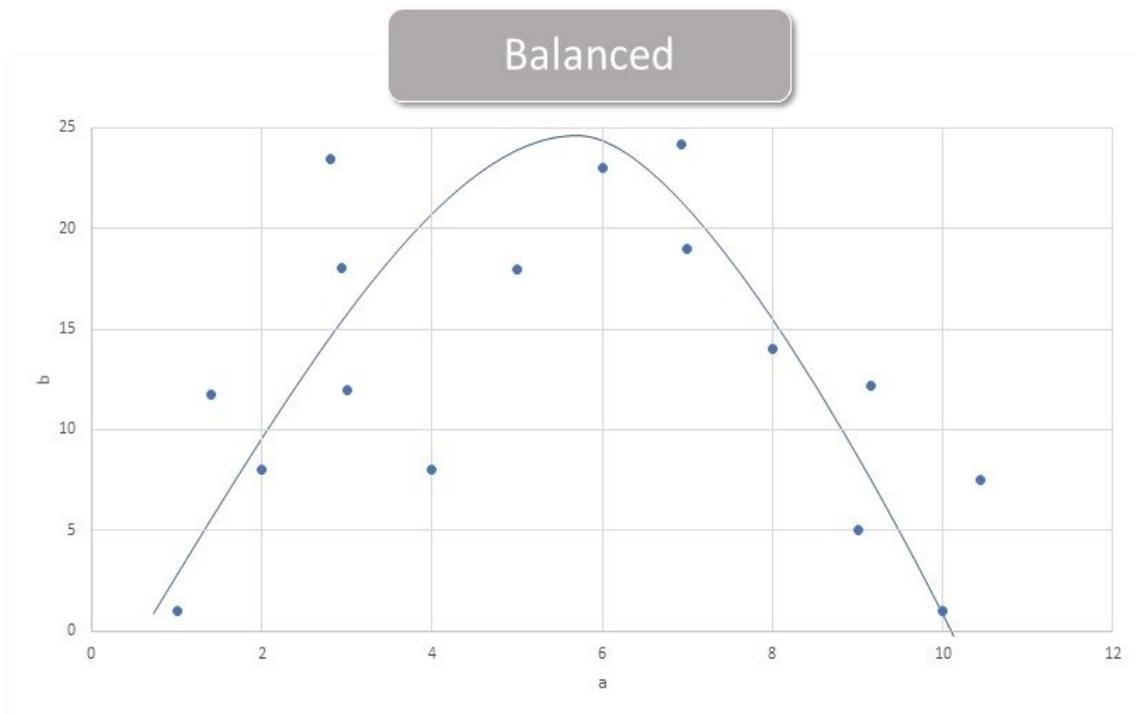


Abbildung 33: *Balanced, Generalisierbarkeit eines Modells*

Datenqualität

In den letzten Jahren wurde der Einsatz von künstlicher Intelligenz (KI), nicht zuletzt für medizinische Anwendungen, enorm gesteigert. Diese neuen technologischen Entwicklungen haben eine ganz neue Diskussion darüber ausgelöst, wie wir große Datenmengen im Gesundheitswesen für die Verarbeitung durch computergestützte Systeme nutzen können. Gerade die Anwendungen im Gesundheitswesen erfordern jedoch ein hohes Maß an (Patienten-)Datenschutz und -Sicherheit. Darüber hinaus wird zunehmend die Forderung nach einer angemessenen Einwilligung des Patienten durchgesetzt (Rosenstein, Capala et al. 2016), was zu zusätzlichen Herausforderungen, insbesondere bei der Erhebung von retrospektiven Daten führt. Eine weitere Schwierigkeit ist, dass viele Daten unstrukturiert und nicht standardisiert sind. Daraus ergeben sich beispielsweise Fragen, welche Daten personenbezogen sind und wem die Daten, die von selbstlernenden Computersystemen generiert werden, tatsächlich gehören. Sind diese Fragen und Probleme erst einmal gelöst, können die Daten gesammelt und genutzt werden. Im Zeitalter des maschinellen Lernens hat das alte

Sprichwort der Informatik „garbage in, garbage out“ wieder an Bedeutung gewonnen (Abbildung 34).

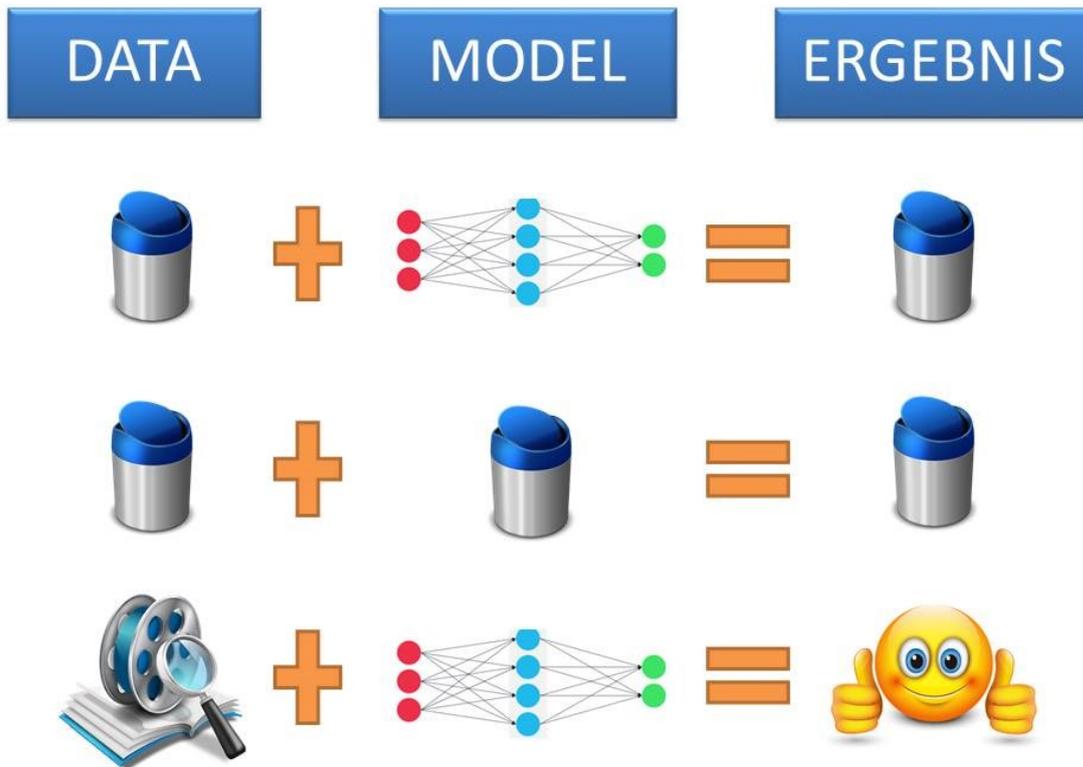


Abbildung 34: Datenqualität

Die Datenqualität ist ein sehr wichtiges Thema, da die Entwicklung des Algorithmus direkt mit der (Qualität der) verwendeten Datensammlung verbunden ist. Oftmals sind die von KI- Systemen gelieferten Ergebnisse vorläufig, da jedes neue Datenbit in das Lernsystem eingegeben werden kann und dadurch sich der Algorithmus potenziell verändert. Daher müssen die Daten im Laufe der Zeit auch eine gleichbleibend hohe Qualität aufweisen, um eine Verschlechterung des Algorithmus durch neu eintreffende Daten und Erkenntnisse zu vermeiden. Dies erfordert nicht nur die Qualität der Datenerhebung, sondern auch einen Prozess der digitalen Kuratierung der gesammelten Daten, um den Wert und die Nutzbarkeit zu erhöhen. Die Datenkuratierung ermöglicht das Auffinden von Daten, erhält ihre Qualität, steigert den Wert und sorgt für die Wiederverwendung im Laufe der Zeit. Dieser neue Bereich umfasst die Authentifizierung, Archivierung, Verwaltung, und Bewahrung von Daten. Dieser Datenkuratierungsprozess gilt als Voraussetzung, um eine auffindbare und

wiederverwendbare bildgebende Datenbank aufzubauen (Prior, Smith et al. 2017). Aktuelle Schätzungen gehen davon aus, dass sich die Gesamtdatenmenge, inklusive medizinischer Daten, auf der Welt alle 2-3 Jahre verdoppelt (Mayer-Schönberger and Ingelsson 2018). Gleichzeitig ist der Anteil der digital erfassten Daten in den letzten zwei Jahrzehnten enorm gestiegen. Die Digitalisierung der Bildgebung begann mit der Einführung einer standardisierten Datenstruktur und Kommunikation mit DICOM (Digital Imaging and Communication in Medicine) und der Entwicklung von Bildarchivierungs- und Kommunikationssystemen (PACS). Diese ermöglichen eine komfortablere und standardisierte Erfassung der Bilddaten und können auch die langfristige Speicherung und Zugänglichkeit der Bilddaten garantieren, sofern geeignete Speichermedien und eine geeignete Migrationsstrategie eingesetzt werden (van Ooijen, Viddeleer et al. 2010). Der Datenzuwachs selbst wurde durch den ständig wachsenden Bedarf an qualitativ hochwertigen Bilddaten ausgelöst und vor allem durch die Entwicklungen in der Computertomographie (CT) und Magnetresonanztomographie (MRT) vorangetrieben. Auch nachbearbeitete Daten können im PACS aufgenommen werden. Bei der Durchführung der retrospektiven Datenerfassung aus einer PACS-Umgebung besteht die Herausforderung darin, aus dem erfassten und generierten Datensatz die entsprechende Auswahl zu treffen, die für die Entwicklung von KI-Algorithmen notwendig ist. Diese Auswahl ist oftmals eine menschliche Entscheidung und damit anfällig für Fehler, Variationen und persönliche oder institutionelle Präferenzen.

Klinische Entscheidungsfindung

Bei der Behandlung eines Patienten sollten idealerweise alle klinischen Angaben für das weitere Management in Betracht gezogen werden. Hierbei könnten uns in naher Zukunft KI-Systeme unterstützend zur Seite stehen. Folgendes, hypothetisches Szenario wäre denkbar:

Ein älterer Patient stellt sich in der Notaufnahme bei Zustand nach Cholezystektomie vor 3 Wochen, mit aktueller Dyspnoe und moderater Tachykardie vor. Nachdem eine Pneumonie sowohl klinisch als auch ggf. im Röntgen-Thorax ausgeschlossen wird, kommt eine mögliche Lungenarterienembolie in Frage. Um die Wahrscheinlichkeit einer Lungenarterienembolie abzuschätzen wird der Wells-Score errechnet. Hiernach fällt dieser Patient in die Gruppe mit mittlerem Risiko (1,5 Punkte für eine Herzfrequenz >100/min, 1,5 Punkte für eine Operation innerhalb der letzten 4 Wochen) (Wells, Anderson et al. 2001). Um eine Lungenarterienembolie zu bestätigen bzw. auszuschließen wird anschließend eine CT-Angiographie durchgeführt. Diese zeigt eine

Kontrastmittelaussparung in der rechten Unterlappenarterie. Somit wird der klinische und anamnestische Verdacht einer Lungenarterienembolie bestätigt. Zudem zeigt die CT-Thorax-Untersuchung eine ca. 7 mm messende, nodulär konfigurierte Läsion im Lungensegment 8 rechts. Nach Fleischner-Kriterien wird eine Verlaufsuntersuchung in 6-12 Monaten empfohlen (MacMahon, Naidich et al. 2017).

Allein in diesem eher einfachen Beispiel gibt es verschiedene Möglichkeiten, wie KI-Systeme den radiologischen Arbeitsablauf erleichtern und zur Unterstützung des Patientenmanagements eingesetzt werden können. Zum Beispiel könnte die bestellte CT-Angiographie auf der Arbeitsliste des Radiologen priorisiert werden, da die Vortestwahrscheinlichkeit für eine Lungenembolie bei diesem Patienten relativ hoch ist (Wells, Anderson et al. 2001). Darüber hinaus könnte ein spezielles KI-System die CT-Studie vorklassifizieren und einen vorläufigen strukturierten Bericht erstellen. Der Radiologe würde in dieser Konstellation als Kontroll- und Validierungsinstanz fungieren. Der Algorithmus würde in Zusammenschau mit der Klinik und der Anamnese (bspw. Familien- und Raucheranamnese) eine leitlinienkonforme Verlaufsuntersuchung bezüglich des Rundherds im rechten Unterlappen empfehlen. Gerade im Hinblick auf die Nachsorge von Nebenbefunden hat sich gezeigt, dass nur ein Drittel aller radiologischer Empfehlungen mit den jeweiligen Leitlinien übereinstimmen (Lacson, Prevedello et al. 2012).

Es ist realistisch anzunehmen, dass sehr bald KI-Anwendungen, dem Radiologen unterstützend beistehen werden. Die automatisierte Analyse von Röntgenthoraxbildern könnte dazu genutzt werden, Patienten mit auffälligen Befunden zu priorisieren. Die Erkennung, Vermessung und Charakterisierung eines Lungenrundherds ist ein weiteres Beispiel für eine solche klar definierte Aufgabe. Neue KI-basierte Anwendungen bieten die Möglichkeit, Auffälligkeiten in Notfallsituationen wie z.B. Schlaganfall schneller zu erkennen. Bei Patienten, die sich zum Ausschluss eines Schlaganfalls einem nativen Schädel-CT unterziehen, ist es für das weitere Procedere entscheidend, die ischämischen Veränderungen im Gehirn so schnell wie möglich zu erkennen. Der e-ASPECTS-Algorithmus ist Neuroradiologen bei der Zuweisung des Alberta Stroke Program Early Computed Tomography Score (ASPECTS) nicht unterlegen (Nagel, Sinha et al. 2017). Viz.ai ist eine von der FDA zugelassene Plattform, die proximale intrakranielle Verschlüsse großer Gefäße in den CT-Angiographie-Bildern (CTA) der Hirngefäße erkennt. Der Algorithmus benachrichtigt daraufhin umgehend den Neuroradiologen über den Befund und beschleunigt damit den Entscheidungsprozess bezüglich der Notwendigkeit einer Thrombektomie und (falls erforderlich) der Überweisung des Patienten an ein spezialisiertes Behandlungszentrum. Ein weiterer vielversprechender, klinischer Anwendungsfall scheint die automatisierte Berechnung

der Fraktionierten Flussreserve (FFR) aus den kardialen CT-Angiographie-Bildern (FFRCT) zu sein. Die CTA der Koronargefäße ist eine gut etablierte Methode zur Abschätzung des Stenosegrades der Koronargefäße. Gemäß den FFR-Messungen sind jedoch oftmals diese Stenosegrad-Messungen nicht von hämodynamischer Bedeutung. Ein Algorithmus kann zur Bestimmung der virtuellen hämodynamischen Bedeutung von Stenosegraden eingesetzt werden und bietet somit eine gute Präselektionsmöglichkeit. Der Einsatz dieser Methode könnte zum Nutzen des Patienten durch die Vermeidung eines nicht notwendigen interventionellen Eingriffs sein.

Der erste Schritt einer sinnvollen und erfolgreichen Implementierung von KI-Anwendungen ist daher die Identifizierung relevanter klinischer Anwendungsfälle, unter anderem durch Radiologen. Nach Identifizierung geeigneter Anwendungsfälle sollte eine klare Definition des Ziels der Anwendung festgelegt werden. Das „Data Science Institute American College of Radiology“ (ACR DSI) spezifiziert derzeit detaillierte Anwendungsfälle für KI-Algorithmen (Allen and Dreyer 2018). Die Einführung von KI-Applikationen zur Übernahme einfacher oder langwieriger Aufgaben, zur Reduzierung der radiologischen Arbeitsbelastung, zur Erleichterung der Analyse von Notfalluntersuchungen und zur Optimierung der Patientenplanung ist technisch nicht weit entfernt. Darüber hinaus wird es dank dieser KI-Anwendungen einfacher sein, Patienten, die an Orten leben, an denen ein Mangel an Radiologen herrscht, wie z.B. in abgelegenen Gebieten und in Entwicklungsländern, eine medizinische Grundversorgung und Diagnostik zu ermöglichen. Ein gutes Beispiel für einen solchen klinischen Anwendungsfall ist die CAD4TB (Computer Aided Detection for Tuberculosis)-Software, die für das Screening von Tuberkulose (TB) auf konventionellen Röntgenthoraxaufnahmen in Gebieten wie zum Beispiel Ghana und Pakistan, eingesetzt wird. Eine retrospektive Auswertung der Software belegte die hohe Sensitivität und Spezifität des Algorithmus (Melendez, Hogeweg et al. 2018).

Erwartungen an die KI

Im Jahr 2017 wurde das Gesundheitswesen als ein wichtiger Bereich für KI-Startup-Investitionen eingestuft, da das Gesundheitswesen unzählige KI-Applikationsmöglichkeiten bietet. Da die medizinische Bildgebung technologisch sehr schnell voranschreitet, sind die Erwartungen an die Radiologie sehr hoch und haben sich zu einem wahren „Hype“ entwickelt. Das Ausmaß wurde auf den Jahrestagungen der amerikanischen und europäischen radiologischen Gesellschaften, der RSNA 2017 in

Chicago und dem ECR 2018 in Wien bemerkbar. Hier herrschte ein großes Interesse an allen KI-bezogenen Themen. Eine entscheidende Frage, die viele Radiologen weltweit beschäftigt, ist, welche Erwartungen an die KI realistisch sind und welche Auswirkungen diese neue Technologie auf die radiologische Fachwelt haben wird. Der Hype-Zyklus nach Gartner ist eine bekannte Darstellung, welche die Entwicklung neuer Technologien und die damit verbundenen Erwartungen darstellt (Abbildung 35). Ähnliche Beispiele für solche „Hype-Zyklen“ sind mit der Einführung des 3D-Drucks in der Radiologie vor einigen Jahren und auch bei den selbstfahrenden Autos zu sehen. Auf den „Gipfel der überzogenen Erwartungen“ folgt in der Regel ein „Tal der Enttäuschung“, der dann eintritt, wenn die anfänglichen Erwartungen nicht schnell erfüllt werden. Auf einen „Pfad der Erleuchtung“, eine Zeit, in der echte Fortschritte gemacht werden, folgt meist ein „Plateau der Produktivität“, in dem die Technologie sich stabilisiert und allgemein anerkannt wird. Wenn die Erwartungen zu hoch angesetzt sind, folgt unvermeidlich eine Enttäuschung, auch wenn die grundlegende Technologie solide ist.

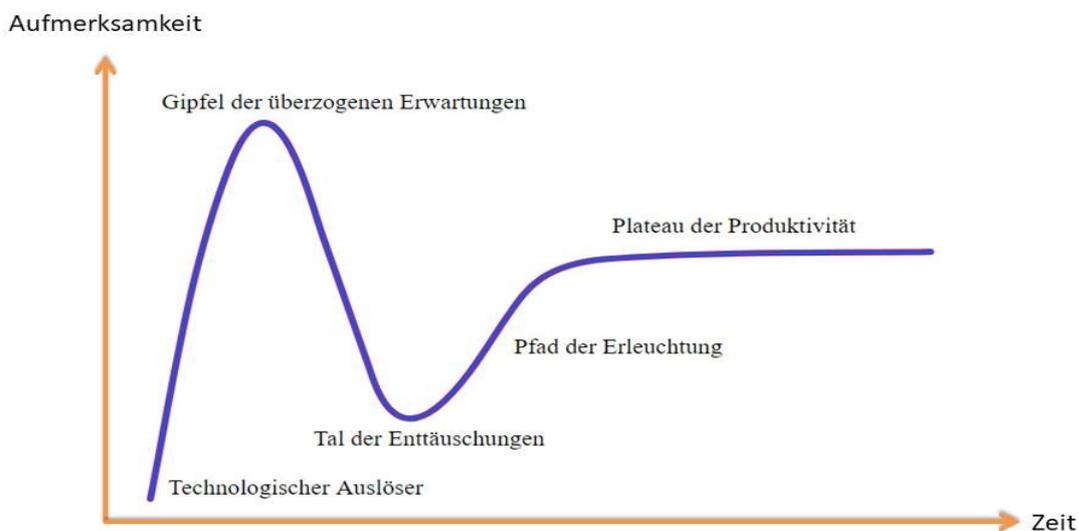


Abbildung 35: Hype-Zyklus nach Gartner

Ausblick und Schlussfolgerung

GO ist ein Brettspiel bei dem zwei Spieler durch abwechselndes Legen von Spielsteinen auf Schnittpunkte der Spielfeldlinien das Ziel haben Territorien zu erobern, diese zu

sichern und nach Möglichkeit Spielsteine des Gegners „gefangen“ zu nehmen. Durch geeignete Formationen können gesicherte Stellungen nach und nach geschaffen werden. Die Größe der eroberten Territorien und die Anzahl der gefangenen Spielsteine bestimmen den Gewinner. GO ist ein komplexes und logisches Spiel. Ein Spieler kann sein Leben lang an der Verfeinerung seines Stils und seiner Spielstärken arbeiten. Die Zahl der spielbaren Varianten ist größer als die Anzahl von Atomen im Universum und ganz deutlich höher als im Spiel Schach. Gespielt wird mit 181 schwarzen und 180 weißen Spielsteinen, die sich untereinander nicht unterscheiden. Meistens werden jedoch deutlich weniger Steine benötigt. Das GO Spielbrett besteht aus einem Raster von horizontalen und vertikalen Linien. Am populärsten ist ein 19x19 Brett mit 361 Schnittpunkten.

In 2015 dachten Experten des Brettspiels GO, dass es aufgrund der Komplexität und Vielfalt der Spielmöglichkeiten noch mindestens 5 bis 10 Jahre dauern würde, ehe Software eine Chance hat gegen GO-Meister erfolgreich anzurechnen. Bereits im März 2016 hat AlphaGO von Google dann den GO-Weltmeister Lee Sedol mit 4 zu 1 geschlagen. Selbst lernende Software generierte eigenständig ganz originelle GO-Züge und verwarf von Menschen gefundene Lehrbuch-Varianten (Silver, Huang et al. 2016). Die Software Version AlphaGo Zero erzeugte im Jahr 2017 neue Strategien, mit denen es gegen Vorversionen gewann, gegen die bereits kein menschlicher Go-Weltmeister mehr gewinnen konnte.

Auch im Brettspiel Atari konnte die künstliche Neuronale Netz Methodik durch ihre Effizienz und Genauigkeit überzeugen (Mnih, Kavukcuoglu et al. 2015). Selbst in der medizinischen Bildgebung beeindruckten Deep Learning Methoden mit ihrer hohen Genauigkeit. In der Pathologie können DNN Adenokarzinome der Lunge auf dem Niveau erfahrener Pathologen in Subtypen klassifizieren (Wei, Tafe et al. 2019). In der Radiologie entwickelten Jung und Mitarbeiter 3D-Neuronale Netze, die Lungenrundherde zuverlässig detektieren (Jung et al. 2018). Ardila und Mitarbeiter entwickelten im Mai 2019 eine Deep Learning Software, die in Lungenkrebsfrüherkennung sechs erfahrene Radiologen in der Genauigkeit übertraf (Ardila, Kiraly et al. 2019). Weitere zahlreiche, aktuelle und hochrangig publizierte Arbeiten zeigen das Potential der Deep Learning Methoden in verschiedenen Bereichen der Medizin (exemplarisch, „Früherkennung der Multiplen Sklerose“, „Automatische Segmentierung der Leitungsbahnen der weißen Substanz“, „Vorhersage von kardiovaskulären Risikofaktoren aus Fundusaufnahmen“, „Klassifizierung von altersbedingter Makuladegeneration und diabetischem Makulaödem“, „Detektion einer pädiatrischen Pneumonie auf Röntgenbildern der Lunge“, „Herausforderungen bei der Anwendung von Deep Learning Algorithmen im klinischen Alltag“, „Maschinelles Lernen-Möglichkeiten für das Gesundheitswesen“,

„Deep Learning zur Früherkennung von Netzhauterkrankungen“ (De Fauw, Ledsam et al. 2018, Ghassemi, Naumann et al. 2018, Hinton 2018, Kermany, Goldbaum et al. 2018, Poplin, Varadarajan et al. 2018, Rieger, Akçakaya et al. 2018, Wasserthal, Neher et al. 2018, Yoo, Tang et al. 2018).

Bei der KI-Software Entwicklung für die Radiologie ist es entsprechend ein nächster Schritt, mit selbstoptimierenden Algorithmen zu experimentieren, die Zugriffe auf große Bild- und Befundarchive haben. Mit entsprechender KI-Software können dann eventuell innerhalb kurzer Zeit Ergebnisse erzielt werden, für die man mit nicht selbst lernender Software Jahrzehnte veranschlagen würde.

Zurzeit zeigen sich Neuronale Netze als effiziente Methode, um bei rein pragmatischen Anwendungen, mit hohen Sensitivitäts- und Spezifitäts-Werten zu beeindrucken. Spektakuläre Klassifikationsergebnisse beruhen dabei auf vorklassifizierten Anlern Daten unter der Anwendung von Rechenoperationen, die noch fern davon entfernt sind, in ihrer Komplexität auch nur annähernd verstanden zu sein. Die Gruppenanzahl, in die klassifiziert wird, ist bei den derzeitigen Publikationen noch sehr klein und besteht oft nur aus benigne oder maligne. Radiologen, Neuroradiologen und Kinderradiologen klassifizieren zwischen hunderten von verschiedenen Erkrankungen inklusive verschiedener Schweregrade. Bis qualifizierte und zertifizierte Lerndaten vorliegen sind noch verschiedene Hürden zu überwinden.

Bevor KI- Anwendungen routinemäßig in der radiologischen Praxis eingesetzt werden, sollte man sich bewusst sein, dass es noch viele bekannte und unbekannte Risiken gibt, die mit dem Einsatz von ML-basierten Lösungen verbunden sind. Obwohl die derzeit bestehenden Schwächen und Probleme dieser Algorithmen nach und nach sichtbar werden, könnten sie für viele zukünftige Anwender aufgrund ihrer mangelnden Grundkenntnisse in dieser Technik möglicherweise verborgen bleiben. Grundkenntnisse in Informatik, Statistik und Deep Learning sollten als Voraussetzung angesehen werden, um zu verstehen und vorauszusehen, was schief gehen kann und wie es vermieden oder verbessert werden sollte. Die physikalischen Grundprinzipien in der medizinischen Bildgebung sind Bestandteil der Ausbildung eines Radiologen, so sollten die Grundprinzipien der KI auch im modernen Ausbildungscurriculum enthalten sein.

Schädlings-Software in der Computerwelt lehrt, dass man nicht jedem Programm vertrauen sollte. Mit absichtlich gefälschten (falsch markierten) Daten kann man Anlern Datensätze manipulieren, sodass Fehlklassifikationen eines Deep Learning Algorithmus irrelevante und irreführende Prognosen liefern. Eine Zertifizierung von Lerndatensätzen scheint daher sinnvoll. Rechtliche Fragen nach dem Verantwortungsträger sind ungeklärt: Wer haftet bei der Fehlentscheidung des Computerprogramms?

Ferner gibt es allgemeine Fragen, die die IT aufwirft, die noch geklärt werden müssen: Mit Computersimulationen (Generative Deep Learning) kann man positiv betrachtet die morphologische Entwicklung von Erkrankungen erkunden, wobei Erkrankungen in ein Bild simuliert werden. Mit entsprechender krimineller Energie können so Patientendaten, aber auch in eine Therapienotwendigkeit hinein manipuliert werden, um lukrative Behandlungskosten zu generieren. Original-Bilddaten würden dann die Grundlage bilden, um den Therapieerfolg zu belegen. Sicherlich werden durch Deep Learning Methoden positive Anwendungsszenarien im Klinikbetrieb und in Lehre und Forschung entstehen. Generell sollte man beachten, dass diese neuen Methoden bislang hauptsächlich pragmatischer Natur sind und nur ein geschulter Kopf mit umfangreichem eigenem Wissen neue Erkenntnisse realisieren kann. Mathematik und Informatik haben noch viele Fragen zu klären, ehe man bewährte Methoden zu schnell ändern sollte.

Zusammenfassung

Das Ziel dieser Arbeit ist es, zu klären, ob eine **automatische Klassifikation von CT-Thoraxbildern** zwischen „Gesund“ und „Fibrose“ **mit Hilfe von Deep Learning Algorithmen** zuverlässig möglich ist.

Hierfür standen 52 HRCT-Datensätze von unauffälligen Thoraxbefunden und 52 Datensätze von Patienten mit fibrotisch erkranktem Lungenparenchym zur Verfügung. Diese Datensätze dienten als Lern-, Validierungs- und Prüfdaten für die neuangelernten Neuronalen Netze. Fünf verschiedene Programmvarianten wurden untersucht.

Variante 1: Zunächst wurden zufällig ausgewählte Bilder aus den Ordnern „Gesund“ bzw. „Fibrose“ ausgewählt und in die entsprechenden Ordner „Lehrdaten_Gesund“ (1351 Schnittbilder), „Lehrdaten_Fibrose“ (942 Schnittbilder), „Prüfdaten_Gesund“ (1169 Schnittbilder) und „Prüfdaten_Fibrose“ (1057 Schnittbilder) kopiert. Mit dieser **bildspezifischen Aufteilung** der Daten wurde das Neuronale Netz **Inception v3** unter **TensorFlow** neu angelernt.

Variante 2: In einer zweiten Programmvariante wurden in die Anlernphase mehr Bilddaten in den Lern- und Prüfungsprozess einbezogen. Es wurden jeweils 26 zufällig gewählte Patienten diagnosespezifisch in die Ordner „Lehrdaten_Gesund“ (5622 Schnittbilder), „Lehrdaten_Fibrose“ (3323 Schnittbilder), „Prüfdaten_Gesund“ (5407 Schnittbilder) und „Prüfdaten_Fibrose“ (2919 Schnittbilder) kopiert. Zudem erfolgte eine **patientenspezifische Aufteilung** der Daten. Mit dieser patientenspezifischen Aufteilung wurde **Inception v3** erneut unter **TensorFlow** angelernt.

Varianten 3 bis 5: Darüber hinaus wurden die Neuronalen Netze **DenseNet121**, **Inception v3** und **ResNet50** unter **PyTorch** angelernt. Hierzu erfolgte eine **patienten- und diagnosespezifische** Aufteilung der Daten in die Ordner „Lerndaten_Gesund“ (23 Patienten), „Validierungsdaten_Gesund“ (3 Patienten), „Prüfdaten_Gesund“ (26 Patienten), „Lerndaten_Fibrose“ (23 Patienten), „Validierungsdaten_Fibrose“ (3 Patienten) und „Prüfdaten_Fibrose“ (26 Patienten). Für einige gewählte Bilder wurden „Class-Activation- Maps“ erstellt, um die Grundlage der Klassifizierungsentscheidung transparenter dazustellen.

Die neu angelernteten Neuronalen Netze konnten erfolgreich eingesetzt werden. Das unter TensorFlow neu angelernte Neuronale Netz Inception v3 (Variante 1) zeigte in der ersten, bildspezifischen Aufteilung der Daten eine Sensitivität von 96,9 % und eine Spezifität von 95,4 %. Die patientenspezifische Aufteilung der Daten (Variante 2) zeigte unter TensorFlow eine Sensitivität von 99,1 % und eine Spezifität von 99,4 %. Unter PyTorch klassifizierten DenseNet121 und Inception v3 alle Prüfdaten richtig (Varianten 3 und 4) und zeigten somit eine Sensitivität und eine Spezifität von 100 %. ResNet50

(Variante 5) klassifizierte unter PyTorch zwei fibrotische Schnittbilder als Gesund (Sensitivität 99,93 %, Spezifität 100 %).

Deep Learning Methoden sind für eine vollautomatische Klassifizierung von CT-Thoraxbildern in „Gesund“ und „Fibrose“ geeignet. Durch weitere Arbeiten soll evaluiert werden, ob eine automatische Fibroseschweregrad- und Mustereinteilung möglich ist.

Summary

Purpose of this work was to clarify the feasibility of using **Deep Learning algorithms in classifying lung CT-scans into healthy and fibrosis**. More detailed, the influence of different AI algorithms with respect to specificity and sensitivity was investigated in a simple classification task between healthy and fibrotic CT lung scans. Additionally, the influence of different selection rules of training, validation and test groups was evaluated. In this study, the dataset consisted of high-resolution CT-scans of 52 patients for each classification group. The models were trained, validated and tested with these data. Five different models were evaluated.

First model: First, randomly selected images were selected from the folders "Gesund" and "Fibrose" and copied to the corresponding folders "Lehrdaten_Gesund" (1351 images), "Lehrdaten_Fibrose" (942 images), "Prüfdaten_Gesund" (1169 images) and "Prüfdaten_Fibrose" (1057 images). With this **image specific distribution** of the data, the neural network **Inception v3** was retrained with **TensorFlow**.

Second model: In a second training process, more image data was included in the learning and examination process. In addition, a **patient-specific division** of the data was performed. Twenty-six randomly selected patients were copied into the folders "Lehrdaten_Gesund" (5622 images), "Lehrdaten_Fibrose" (3323 images), "Prüfdaten_Gesund" (5407 images) and "Prüfdaten_Fibrose" (2919 images). With this patient-specific division, **Inception v3** was retrained with **TensorFlow**.

Third to 5th model: Additionally, the neural networks **DenseNet121**, **Inception v3** and **ResNet50** were retrained using the program library **PyTorch**. For this purpose, **patient- and diagnosis-specific** data were divided into the folders "Lerndaten_Gesund" (23 patients), "Validierungsdaten_Gesund" (3 patients), "Prüfdaten_Gesund" (26 patients), "Lerndaten_Fibrose" (23 patients), "Validierungsdaten_Fibrose" (3 patients) and "Prüfdaten_Fibrose" (26 patients). **Class-Activation-Maps** were generated for some selected images to visualize the basis of the classification decision of the software.

The retrained neural networks were successful in classifying lung CT-scans into healthy and fibrosis. In the first image-specific distribution of the data, the neuronal network Inception v3 (TensorFlow) performed with a sensitivity of 96.9 % and a specificity of 95.4 %. The patient-specific distribution of the data led to a sensitivity of 99.1 % and a specificity of 99.4% (TensorFlow). DenseNet121 (PyTorch) and Inception v3 (PyTorch) correctly classified all test data, showing a sensitivity and a specificity of 100 %. ResNet50 (PyTorch) classified two fibrotic images as healthy (sensitivity 99.93 %, specificity 100 %).

These findings suggest that Deep learning methods are robust and suitable for a fully automated classification of CT thoracic images into "healthy" and "fibrosis". Further studies are required to evaluate an automated image classification performance into different fibrosis severity grades and their corresponding morphological patterns.

Abbildungsverzeichnis

Abbildung 1: Deep Learning als Subklasse	13
Abbildung 2: Supervised Learning als Lernverfahren.....	14
Abbildung 3: Clustering – automatische Gruppenzuordnung	14
Abbildung 4: vereinfachte Darstellung eines Neurons.....	15
Abbildung 5: einfaches vs. tiefgehendes Neuronales Netz	17
Abbildung 6: Lernansätze – Training from Scratch vs. Transfer Learning. Bei Transfer Learning werden vorangelernte Netzwerke, die beispielsweise zwischen Hunden und Katzen differenzieren können mit neuen Daten neu angeleert.....	18
Abbildung 7: Aufteilungsprinzip der Daten	18
Abbildung 8: Datenstruktur für TensorFlow.....	23
Abbildung 9: Datenstruktur für PyTorch	26
Abbildung 10: Beispiel einer Class-Activation-Map	30
Abbildung 11: richtig klassifiziert (TensorFlow, Inception v3, Variante 1)	33
Abbildung 12: falsch klassifiziert (TensorFlow, Inception v3, Variante 1)	34
Abbildung 13: falsch klassifiziert (TensorFlow, Inception v3, Variante 2)	35
Abbildung 14: richtig klassifiziert, Zuordnungswahrscheinlichkeit 1 (TensorFlow, Inception v3, Variante 2).....	36
Abbildung 15: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,9 - 0,99 (TensorFlow, Inception v3, Variante 2)	37
Abbildung 16: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,8 - 0,89 (TensorFlow, Inception v3, Variante 2)	38
Abbildung 17: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,7 - 0,79 (TensorFlow, Inception v3, Variante 2)	39
Abbildung 18: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,6 - 0,69 (TensorFlow, Inception v3, Variante 2)	40
Abbildung 19: richtig klassifiziert, Zuordnungswahrscheinlichkeit 0,5 - 0,59 (TensorFlow, Inception v3, Variante 2)	41
Abbildung 20: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, DenseNet121)	42
Abbildung 21: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, Inception v3)	43
Abbildung 22: richtig klassifiziert mit zugehörigen Heat Maps (PyTorch, ResNet50)	44
Abbildung 23: falsch klassifiziert mit zugehörigen Heat Maps (PyTorch, ResNet50)	45
Abbildung 24: Lern- und Validierungsgenauigkeit während der Anlernphase - DenseNet121	46

Abbildung 25: Lern- und Validierungsgenauigkeit während der Anlernphase - Inception v3.....	46
Abbildung 26: Lern- und Validierungsgenauigkeit während der Anlernphase - ResNet50	47
Abbildung 27: Explainable AI.....	54
Abbildung 28: regelbasierte Methoden-Eine Aufgabe wird mittels mathematisch definierter Regeln gelöst. Ist das Ergebnis zufriedenstellend, kann das Modell in Anwendung gebracht werden. Sollte das Ergebnis nicht zufriedenstellend sein, sollten die zuvor mathematisch definierten Regeln überdacht bzw. geändert werden.....	55
Abbildung 29: maschinelles Lernen-Ein Modell wird anhand Anlern Daten und einem bestimmten Anlernalgorithmus im Hinblick auf eine bestimmte Fragestellung angelernt. Ist das Ergebnis des Modells zufriedenstellend, kann es in Anwendung gebracht werden. Ist das Ergebnis nicht zufriedenstellend, sollten die Anlern Daten auf eine ausreichende Quantität/Qualität bzw. der Anlernalgorithmus auf Fehler überprüft werden.	56
Abbildung 30: Daten besser verstehen-Ein Modell wird anhand Anlern Daten und einem bestimmten Anlernalgorithmus im Hinblick auf eine bestimmte Fragestellung angelernt. Ist das Ergebnis des Modells zufriedenstellend, kann der Lösungsweg dargestellt werden. Dies kann zum besseren Verständnis der Daten beitragen. Ist das Ergebnis nicht zufriedenstellend, sollten die Anlern Daten bzw. der Anlernalgorithmus auf suboptimale Lösungsansätze überprüft werden.....	56
Abbildung 31: Overfitting-gefittete Linie trifft exakt alle Punkte	62
Abbildung 32: Underfitting-gefittete Linie korreliert weder mit Parameter A oder B ...	62
Abbildung 33: Balanced, Generalisierbarkeit eines Modells.....	63
Abbildung 34: Datenqualität.....	64
Abbildung 35: Hype-Zyklus nach Gartner	68

Tabellenverzeichnis

Tabelle 1: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (TensorFlow, Inception v3, Variante 1)	48
Tabelle 2: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (TensorFlow, Inception v3, Variante 2)	49
Tabelle 3: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, DenseNet121)	50
Tabelle 4: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, Inception v3)	51
Tabelle 5: Anzahl der korrekt zugeordneten Bilder, Spezifität und Sensitivität für gesundes und fibrotisches Lungenparenchym gruppiert nach Klassifizierungswahrscheinlichkeitsbereichen (PyTorch, ResNet50)	52

Literaturverzeichnis

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467 [cs]*.
- Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., . . . Rietveld, D. J. N. c. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *5*(1), 1-9.
- Allen, B., & Dreyer, K. J. J. o. t. A. C. o. R. (2018). The artificial intelligence ecosystem for the radiological sciences: ideas to clinical practice. *15*(10), 1455-1457.
- Anavi, Y., Kogan, I., Gelbart, E., Geva, O., & Greenspan, H. (2016). *Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval*.
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, *35*(5), 1207-1216. doi:10.1109/TMI.2016.2535865
- Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., . . . Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954-961. doi:10.1038/s41591-019-0447-x
- Armanios, M. Y., Chen, J. J.-L., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., . . . Phillips III, J. A. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. *New England Journal of Medicine*, *356*(13), 1317-1326.
- Avanzo, M., Stancanello, J., & El Naqa, I. J. P. M. (2017). Beyond imaging: the promise of radiomics. *38*, 122-139.
- Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2018). Chest pathology identification using deep feature selection with non-medical training. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *6*(3), 259-263. doi:10.1080/21681163.2016.1138324
- Baumgartner, K. B., Samet, J. M., Stidley, C. A., Colby, T. V., & Waldron, J. A. (1997). Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, *155*(1), 242-248.
- Bhooshan, N., Giger, M. L., Jansen, S. A., Li, H., Lan, L., & Newstead, G. M. J. R. (2010). Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers. *254*(3), 680-690.
- Briot, J.-P., Hadjeres, G., & Pachet, F. (2017). Deep learning techniques for music generation-a survey. *arXiv preprint arXiv:1709.01620*.
- Cai, T., Giannopoulos, A. A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K. K., . . . Mitsouras, D. J. R. (2016). Natural language processing technologies in radiology research and clinical applications. *36*(1), 176-191.
- Cameron, A., Khalvati, F., Haider, M. A., & Wong, A. J. I. T. o. B. E. (2015). MAPS: a quantitative radiomics approach for prostate cancer detection. *63*(6), 1145-1156.
- Carneiro, G., Oakden-Rayner, L., Bradley, A., Nascimento, J., & Palmer, L. (2016). *Automated 5-year Mortality Prediction using Deep Learning and Radiomics Features from Chest Computed Tomography*.
- Channin, D. S., Mongkolwat, P., Kleper, V., & Rubin, D. L. (2009). The annotation and image mark-up project. In: Radiological Society of North America, Inc.
- Charbonnier, J. P., Rikxoort, E. M. V., Setio, A. A. A., Schaefer-Prokop, C. M., Ginneken, B. V., & Ciompi, F. (2017). Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Med Image Anal*, *36*, 52-60. doi:10.1016/j.media.2016.11.001

- Chen, W., Giger, M. L., & Bick, U. J. A. r. (2006). A fuzzy c-means (fcm)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced mr images1. *13*(1), 63-72.
- Christodoulidis, S., Anthimopoulos, M., Ebner, L., Christe, A., & Mougiakakou, S. (2017). Multisource Transfer Learning With Convolutional Neural Networks for Lung Pattern Analysis. *IEEE J Biomed Health Inform*, *21*(1), 76-84. doi:10.1109/JBHI.2016.2636929
- Cicero, M., Bilbily, A., Colak, E., Dowdell, T., Gray, B., Perampaladas, K., & Barfett, J. (2017). Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. *Invest Radiol*, *52*(5), 281-287. doi:10.1097/RLI.0000000000000341
- Ciampi, F., Chung, K., van Riel, S. J., Setio, A. A. A., Gerke, P. K., Jacobs, C., . . . van Ginneken, B. (2017). Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*, *7*, 46479. doi:10.1038/srep46479
- Ciampi, F., de Hoop, B., van Riel, S. J., Chung, K., Scholten, E. T., Oudkerk, M., . . . van Ginneken, B. (2015). Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal*, *26*(1), 195-202. doi:10.1016/j.media.2015.08.001
- Coroller, T. P., Grossmann, P., Hou, Y., Velazquez, E. R., Leijenaar, R. T., Hermann, G., . . . Oncology. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *114*(3), 345-350.
- Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., . . . Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine*, *24*(10), 1559-1567. doi:10.1038/s41591-018-0177-5
- Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2011). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing*, *20*(1), 30-42.
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., . . . Visentin, D. J. N. m. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *24*(9), 1342.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F. F. (2009). *ImageNet: a Large-Scale Hierarchical Image Database*.
- Deo Rahul, C. (2015). Machine Learning in Medicine. *Circulation*, *132*(20), 1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593
- Dhungel, N., Carneiro, G., & Bradley, A. P. J. M. i. a. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *37*, 114-128.
- Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*: Springer International Publishing.
- Ding, J., Li, A., Hu, Z., & Wang, L. (2017). Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks. *arXiv e-prints*. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170604303D>
- Doi, K. J. C. m. i., & graphics. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *31*(4-5), 198-211.
- Dou, Q., Chen, H., Yu, L., Qin, J., & Heng, P. A. (2017). Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Trans Biomed Eng*, *64*(7), 1558-1567. doi:10.1109/TBME.2016.2613502
- Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., & Heng, P. A. (2017). 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal*, *41*, 40-54. doi:10.1016/j.media.2017.05.001
- Dvorak, P., & Menze, B. (2015). *Structured prediction with convolutional neural networks for multimodal brain tumor segmentation*.

- Earnest 4th, F., Kelly, P., Scheithauer, B., Kall, B., Cascino, T., Ehman, R. L., . . . Axley, P. J. R. (1988). Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy. *166*(3), 823-827.
- El Naqa, I., Grigsby, P., Apte, A., Kidd, E., Donnelly, E., Khullar, D., . . . Laforest, R. J. P. r. (2009). Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *42*(6), 1162-1171.
- El Naqa, I., Li, R., & Murphy, M. J. J. T. A. (2015). Machine learning in radiation oncology. 57-70.
- El Naqa, I. J. C., & imaging, t. (2014). The role of quantitative PET in predicting cancer treatment outcomes. *2*(4), 305-320.
- Elter, M., & Horsch, A. J. M. p. (2009). CADx of mammographic masses and clustered microcalcifications: a review. *36*(6Part1), 2052-2068.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*, 115. doi:10.1038/nature21056
- Fabris, F., Magalhães, J. P. d., & Freitas, A. A. (2017). A review of supervised machine learning applied to ageing research. *Biogerontology*, *18*(2), 171-188. doi:10.1007/s10522-017-9683-y
- Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1915-1929. doi:10.1109/TPAMI.2012.231
- Flaherty, K. R., Andrei, A.-C., King, T. E., Jr., Raghu, G., Colby, T. V., Wells, A., . . . Martinez, F. J. (2007). Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *American journal of respiratory and critical care medicine*, *175*(10), 1054-1060. doi:10.1164/rccm.200606-833OC
- Flaherty, K. R., King, T. E., Jr., Raghu, G., Lynch, J. P., 3rd, Colby, T. V., Travis, W. D., . . . Martinez, F. J. (2004). Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med*, *170*(8), 904-910. doi:10.1164/rccm.200402-147OC
- Ganeshan, B., Abaleke, S., Young, R. C., Chatwin, C. R., & Miles, K. A. J. C. i. (2010). Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *10*(1), 137.
- Ganeshan, B., Goh, V., Mandeville, H. C., Ng, Q. S., Hoskin, P. J., & Miles, K. A. J. R. (2013). Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *266*(1), 326-336.
- Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H. C., . . . Mollura, D. J. (2018). Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis*, *6*(1), 1-6. doi:10.1080/21681163.2015.1124249
- Gerstmair, A., Daumke, P., Simon, K., Langer, M., & Kotter, E. J. E. r. (2012). Intelligent image retrieval based on radiology reports. *22*(12), 2750-2758.
- Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., & Ranganath, R. J. a. p. a. (2018). Opportunities in machine learning for healthcare.
- Giesel, F., Schneider, F., Kratochwil, C., Rath, D., Holland-Letz, T., Moltz, J., . . . Flechsig, P. J. J. o. N. M. (2016). CT radiomic analysis using lymph-node-density profile in correlation to SUV-value for PET/CT based N-staging. *jnumed*. 116.179648.
- Giger, M. L., Karssemeijer, N., & Schnabel, J. A. J. A. r. o. b. e. (2013). Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *15*, 327-357.
- Giger, M. L. J. J. o. t. A. C. o. R. (2018). Machine learning in medical imaging. *15*(3), 512-520.
- Gillies, R. J., Kinahan, P. E., & Hricak, H. J. R. (2016). Radiomics: images are more than pictures, they are data. *278*(2), 563-577.
- Gnep, K., Fargeas, A., Gutiérrez-Carvajal, R. E., Commandeur, F., Mathieu, R., Ospina, J. D., . . . Hatt, M. J. J. o. M. R. I. (2017). Haralick textural features on T2-weighted MRI are

- associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *45*(1), 103-117.
- Guo, Y., Gao, Y., & Shen, D. (2016). Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching. *IEEE Trans Med Imaging*, *35*(4), 1077-1089. doi:10.1109/TMI.2015.2508280
- Hajnal, J. V., & Hill, D. L. (2001). *Medical image registration*: CRC press.
- Hansell, D. M., Bankier, A. A., MacMahon, H., McCloud, T. C., Muller, N. L., & Remy, J. (2008). Fleischner Society: glossary of terms for thoracic imaging. *Radiology*, *246*(3), 697-722. doi:10.1148/radiol.2462070712
- Hassanpour, S., & Langlotz, C. P. J. A. i. i. m. (2016). Information extraction from multi-institutional radiology reports. *66*, 29-39.
- Hatt, M., Tixier, F., Pierce, L., Kinahan, P. E., Le Rest, C. C., Visvikis, D. J. E. j. o. n. m., & imaging, m. (2017). Characterization of PET/CT images using texture analysis: the past, the present... any future? , *44*(1), 151-165.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]*.
- Hinton, G. J. J. (2018). Deep learning—a technology with the potential to transform health care. *320*(11), 1101-1102.
- Hopkins, R. B., Burke, N., Fell, C., Dion, G., & Kolb, M. (2016). Epidemiology and survival of idiopathic pulmonary fibrosis from national data in Canada. *European Respiratory Journal*, *48*(1), 187-195.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., . . . Liang, C. J. R. (2016). Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non—small cell lung cancer. *281*(3), 947-957.
- Hunninghake, G. W., Lynch, D. A., Galvin, J. R., Gross, B. H., Muller, N., Schwartz, D. A., . . . Hogg, J. C. (2003). Radiologic findings are strongly associated with a pathologic diagnosis of usual interstitial pneumonia. *Chest*, *124*(4), 1215-1223. doi:10.1378/chest.124.4.1215
- Hutchinson, J., Fogarty, A., Hubbard, R., & McKeever, T. (2015). Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *European Respiratory Journal*, *46*(3), 795-806.
- Hutchinson, J. P., Fogarty, A. W., McKeever, T. M., & Hubbard, R. B. (2016). In-Hospital Mortality after Surgical Lung Biopsy for Interstitial Lung Disease in the United States. 2000 to 2011. *Am J Respir Crit Care Med*, *193*(10), 1161-1167. doi:10.1164/rccm.201508-1632OC
- Hwang, E. J., Park, S., Jin, K.-N., Kim, J. I., Choi, S. Y., Lee, J. H., . . . Evaluation, G. (2018). Development and Validation of a Deep Learning-Based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clinical Infectious Diseases*, *ciy967-ciy967*. doi:10.1093/cid/ciy967
- Jain, A. K., Jianchang, M., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, *29*(3), 31-44. doi:10.1109/2.485891
- Jennings, H. How is the Nervous System Organized? In (Vol. 2018).
- Jung, H., Kim, B., Lee, I., Lee, J., & Kang, J. (2018). Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. *BMC Med Imaging*, *18*(1), 48. doi:10.1186/s12880-018-0286-0

- Karlo, C. A., Di Paolo, P. L., Chaim, J., Hakimi, A. A., Ostrovnaya, I., Russo, P., . . . Akin, O. J. R. (2014). Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. *270*(2), 464-471.
- Kermary, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., . . . Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, *172*(5), 1122-1131. e1129. doi:10.1016/j.cell.2018.02.010
- Khalvati, F., Wong, A., & Haider, M. A. J. B. m. i. (2015). Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *15*(1), 27.
- King, T. E., Jr., Tooze, J. A., Schwarz, M. I., Brown, K. R., & Cherniack, R. M. (2001). Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med*, *164*(7), 1171-1181. doi:10.1164/ajrccm.164.7.2003140
- Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., . . . Botsis, T. J. J. o. b. i. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *73*, 14-29.
- Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2012). *ImageNet Classification with Deep Convolutional Neural Networks* (Vol. 25).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in neural information processing systems.
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., . . . Fenstermacher, D. J. M. r. i. (2012). Radiomics: the process and the challenges. *30*(9), 1234-1248.
- Kuo, M. D., Gollub, J., Sirlin, C. B., Ooi, C., Chen, X. J. J. o. V., & Radiology, I. (2007). Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. *18*(7), 821-830.
- Lacson, R., Prevedello, L. M., Andriole, K. P., Gill, R., Lenoci-Edwards, J., Roy, C., . . . Khorasani, R. J. J. o. t. A. C. o. R. (2012). Factors associated with radiologists' adherence to Fleischner Society guidelines for management of pulmonary nodules. *9*(7), 468-473.
- Lakhani, P., Prater, A. B., Hutson, R. K., Andriole, K. P., Dreyer, K. J., Morey, J., . . . Itri, J. N. J. J. o. t. A. C. o. R. (2018). Machine learning in radiology: applications beyond image interpretation. *15*(2), 350-359.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., . . . Dekker, A. J. E. j. o. c. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *48*(4), 441-446.
- Le, Q. V. (2013, 26-31 May 2013). *Building high-level features using large scale unsupervised learning*. Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436. doi:10.1038/nature14539
- Ley, B., Collard, H. R., & King Jr, T. E. (2011). Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, *183*(4), 431-440.
- Li, H., Zhu, Y., Burnside, E. S., Drukker, K., Hoadley, K. A., Fan, C., . . . Net, J. M. J. R. (2016). MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *281*(2), 382-391.
- Li, H., Zhu, Y., Burnside, E. S., Huang, E., Drukker, K., Hoadley, K. A., . . . Net, J. M. J. N. b. c. (2016). Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *2*(1), 1-10.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. doi:<https://doi.org/10.1016/j.media.2017.07.005>

- Lo, S. B., Lou, S. A., Lin, J. S., Freedman, M. T., Chien, M. V., & Mun, S. K. (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Trans Med Imaging*, *14*(4), 711-718. doi:10.1109/42.476112
- MacMahon, H., Naidich, D. P., Goo, J. M., Lee, K. S., Leung, A. N., Mayo, J. R., . . . Prokop, M. J. R. (2017). Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *284*(1), 228-243.
- Mayer-Schönberger, V., & Ingelsson, E. J. J. o. i. m. (2018). Big Data and medicine: a big deal? , *283*(5), 418-429.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133. doi:10.1007/BF02478259
- Melendez, J., Hogeweg, L., Sánchez, C., Philipsen, R., Aldridge, R., Hayward, A., . . . Disease, L. (2018). Accuracy of an automated system for tuberculosis detection on chest radiographs in high-risk screening. *22*(5), 567-571.
- Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J. (2011, 11-15 Dec. 2011). *Strategies for training large scale neural network language models*. Paper presented at the 2011 IEEE Workshop on Automatic Speech Recognition & Understanding.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529. doi:10.1038/nature14236
- Mocanu, D. C., Mocanu, E., Stone, P., Nguyen, P. H., Gibescu, M., & Liotta, A. (2018). Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nature Communications*, *9*(1), 2383. doi:10.1038/s41467-018-04316-3
- Moeskops, P., Wolterink, J. M., van der Velden, B. H. M., Gilhuijs, K. G. A., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Deep Learning for Multi-Task Medical Image Segmentation in Multiple Modalities. *arXiv e-prints*. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170403379M>
- Nagel, S., Sinha, D., Day, D., Reith, W., Chapot, R., Papanagiotou, P., . . . Fassbender, K. J. I. J. o. S. (2017). e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. *12*(6), 615-622.
- Network, I. P. F. C. R. (2012). Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *New England Journal of Medicine*, *366*(21), 1968-1977.
- Ni, Y., Wright, J., Perentesis, J., Lingren, T., Deleger, L., Kaiser, M., . . . making, d. (2015). Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *15*(1), 28.
- Nogee, L. M., Dunbar, A. E., 3rd, Wert, S. E., Askin, F., Hamvas, A., & Whitsett, J. A. (2001). A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med*, *344*(8), 573-579. doi:10.1056/nejm200102223440805
- Noth, I., Zhang, Y., Ma, S. F., Flores, C., Barber, M., Huang, Y., . . . Garcia, J. G. N. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med*, *1*(4), 309-317. doi:10.1016/s2213-2600(13)70045-6
- Oakden-Rayner, L. J. U. h. l. w. c. c.-i.-d.-r. (2018). CheXNet: an in-depth review.
- Obert, M. (2019). Are estimations of radiomic image markers dispensable due to recent deep learning findings? *European Respiratory Journal*. doi:10.1183/13993003.01185-2019

- Panth, K. M., Leijenaar, R. T., Carvalho, S., Lieuwes, N. G., Yaromina, A., Dubois, L., . . . Oncology. (2015). Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. *116*(3), 462-466.
- Parekh, V., Jacobs, M. A. J. E. r. o. p. m., & development, d. (2016). Radiomics: a new application from established techniques. *1*(2), 207-226.
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P., & Aerts, H. J. J. S. r. (2015). Machine learning methods for quantitative radiomic biomarkers. *5*, 13087.
- Parmar, C., Grossmann, P., Rietveld, D., Rietbergen, M. M., Lambin, P., & Aerts, H. J. J. F. i. o. (2015). Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *5*, 272.
- Parmar, C., Leijenaar, R. T., Grossmann, P., Velazquez, E. R., Bussink, J., Rietveld, D., . . . Aerts, H. J. J. S. r. (2015). Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *5*, 11044.
- Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., . . . Webster, D. R. J. N. B. E. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *2*(3), 158.
- Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R. K., . . . Wiest, R. J. P. o. (2014). Multi-modal glioblastoma segmentation: man versus machine. *9*(5).
- Prior, F., Smith, K., Sharma, A., Kirby, J., Tarbox, L., Clark, K., . . . Freymann, J. J. S. d. (2017). The public cancer radiology imaging collections of The Cancer Imaging Archive. *4*, 170124.
- Qi, D., Hao, C., Lequan, Y., Lei, Z., Jing, Q., Defeng, W., . . . Pheng-Ann, H. (2016). Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks. *IEEE Trans Med Imaging*, *35*(5), 1182-1195. doi:10.1109/TMI.2016.2528129
- Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., . . . Lasky, J. A. (2011). An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American journal of respiratory and critical care medicine*, *183*(6), 788-824.
- Raghu, G., Weycker, D., Edelsberg, J., Bradford, W. Z., & Oster, G. (2006). Incidence and prevalence of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, *174*(7), 810-816.
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., . . . Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, *1*(1), 18. doi:10.1038/s41746-018-0029-1
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., . . . Shpanskaya, K. J. a. p. a. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning.
- Ranschaert, E. R., Morozov, S., & Algra, P. R. (2019). *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*: Springer.
- Richeldi, L., Collard, H. R., & Jones, M. G. (2017). Idiopathic pulmonary fibrosis. *The Lancet*, *389*(10082), 1941-1952. doi:10.1016/S0140-6736(17)30866-8
- Rieger, B., Akçakaya, M., Pariente, J. C., Llufrui, S., Martinez-Heras, E., Weingärtner, S., & Schad, L. R. (2018). Time efficient whole-brain coverage with MR Fingerprinting using slice-interleaved echo-planar-imaging. *Scientific Reports*, *8*(1), 6667. doi:10.1038/s41598-018-24920-z
- Rosenstein, B. S., Capala, J., Efstathiou, J. A., Hammerbacher, J., Kerns, S. L., Ostrer, H., . . . Xiao, Y. J. I. J. o. R. O. B. P. (2016). How will big data improve clinical and basic research in radiation therapy? , *95*(3), 895-904.
- Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., Mohamed, A.-r., Dahl, G., & Ramabhadran, B. (2015). Deep Convolutional Neural Networks for Large-scale Speech Tasks. *Neural Networks*, *64*, 39-48. doi:<https://doi.org/10.1016/j.neunet.2014.08.005>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117. doi:<https://doi.org/10.1016/j.neunet.2014.09.003>

- Selman, M., King, T. E., & Pardo, A. (2001). Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Annals of internal medicine*, 134(2), 136-151.
- Setio, A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., . . . van Ginneken, B. (2016). Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging*, 35(5), 1160-1169. doi:10.1109/TMI.2016.2536809
- Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015). Multi-scale Convolutional Neural Networks for Lung Nodule Classification. *Inf Process Med Imaging*, 24, 588-599.
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., . . . Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *arXiv e-prints*. <https://ui.adsabs.harvard.edu/#abs/2016arXiv160203409S>
- Sihong, C., Jing, Q., Xing, J., Baiying, L., Tianfu, W., Dong, N., & Jie-Zhi, C. (2017). Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images. *IEEE Trans Med Imaging*, 36(3), 802-814. doi:10.1109/TMI.2016.2629462
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., . . . Lanctot, M. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., . . . Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484. doi:10.1038/nature16961
- Society, A. T. (2000). Idiopathic pulmonary fibrosis: diagnosis and treatment: international consensus statement. *Am J Respir Crit Care Med*, 161, 646-664.
- Stuart, B. D., Choi, J., Zaidi, S., Xing, C., Holohan, B., Chen, R., . . . Garcia, C. K. (2015). Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet*, 47(5), 512-517. doi:10.1038/ng.3278
- Summers, R. M. J. A. r. (2017). Texture analysis in radiology: Does the emperor have no clothes?, 42(2), 342-345.
- Sun, W., Tseng, T. B., Zhang, J., & Qian, W. (2017). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Comput Med Imaging Graph*, 57, 4-9. doi:10.1016/j.compmedimag.2016.07.004
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *arXiv:1512.00567 [cs]*.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Jianming, L. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans Med Imaging*, 35(5), 1299-1312. doi:10.1109/TMI.2016.2535302
- Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2017). *ArtGAN: Artwork synthesis with conditional categorical GANs*. Paper presented at the 2017 IEEE International Conference on Image Processing (ICIP).
- Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. J. I. t. o. i. t. i. b. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. 13(2), 236-251.
- Teramoto, A., Fujita, H., Yamamuro, O., & Tamaki, T. (2016). Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med Phys*, 43(6), 2821-2827. doi:10.1118/1.4948498
- Tesauro, G., Gondek, D. C., Lenchner, J., Fan, J., & Prager, J. M. J. J. o. A. I. R. (2013). Analysis of watson's strategies for playing Jeopardy!, 47, 205-251.

- Tixier, F., Hatt, M., Valla, C., Fleury, V., Lamour, C., Ezzouhri, S., . . . Le Rest, C. C. J. J. o. N. M. (2014). Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *55*(8), 1235-1241.
- Toell, T., Boehme, C., Mayer, L., Krebs, S., Lang, C., Willeit, K., . . . Schoenherr, G. J. B. n. (2018). Pragmatic trial of multifaceted intervention (STROKE-CARD care) to reduce cardiovascular risk and improve quality-of-life after ischaemic stroke and transient ischaemic attack–study protocol. *18*(1), 187.
- Tompson, J., Jain, A., LeCun, Y., & Bregler, C. (2014). Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation.
- Travis, W. D., Costabel, U., Hansell, D. M., King, T. E., Jr., Lynch, D. A., Nicholson, A. G., . . . Valeyre, D. (2013). An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*, *188*(6), 733-748. doi:10.1164/rccm.201308-1483ST
- Tsakiri, K. D., Cronkhite, J. T., Kuan, P. J., Xing, C., Raghu, G., Weissler, J. C., . . . Garcia, C. K. (2007). Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proceedings of the National Academy of Sciences*, *104*(18), 7552-7557.
- Vallières, M., Freeman, C. R., Skamene, S. R., El Naqa, I. J. P. i. M., & Biology. (2015). A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *60*(14), 5471.
- van Ooijen, P. M., Viddeleer, A., Meijer, F., & Oudkerk, M. J. J. o. d. i. (2010). Accessibility of Data Backup on CD-R after 8 to 11 years. *23*(1), 95-99.
- van Tulder, G., & de Bruijne, M. (2016). *Combining Generative and Discriminative Representation Learning for Lung CT Analysis With Convolutional Restricted Boltzmann Machines* (Vol. 35).
- Vázquez, F. (2017). Deep Learning made easy with Deep Cognition. Retrieved from <https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351>
- Wang, H., & Raj, B. (2017). On the Origin of Deep Learning. *arXiv e-prints*. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170207800W>
- Wang, J., Kato, F., Oyama-Manabe, N., Li, R., Cui, Y., Tha, K. K., . . . Shirato, H. J. P. o. (2015). Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: a pilot radiomics study. *10*(11).
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). *Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Wang, Y., Kuan, P. J., Xing, C., Cronkhite, J. T., Torres, F., Rosenblatt, R. L., . . . Garcia, C. K. (2009). Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am J Hum Genet*, *84*(1), 52-59. doi:10.1016/j.ajhg.2008.11.010
- Wasserthal, J., Neher, P., & Maier-Hein, K. H. J. N. (2018). Tractseg-fast and accurate white matter tract segmentation. *183*, 239-253.
- Wei, J. W., Tafe, L. J., Linnik, Y. A., Vaickus, L. J., Tomita, N., & Hassanpour, S. (2019). Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep*, *9*(1), 3358. doi:10.1038/s41598-019-40041-7
- Wells, P. S., Anderson, D. R., Rodger, M., Stiell, I., Dreyer, J. F., Barnes, D., . . . Kovacs, M. J. J. A. o. i. m. (2001). Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *135*(2), 98-107.

- Wu, W., Parmar, C., Grossmann, P., Quackenbush, J., Lambin, P., Bussink, J., . . . Aerts, H. J. J. F. i. o. (2016). Exploratory study to identify radiomics classifiers for lung cancer histology. *6*, 71.
- Yang, W., Chen, Y., Liu, Y., Zhong, L., Qin, G., Lu, Z., . . . Chen, W. (2017). Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Med Image Anal*, *35*, 421-433. doi:10.1016/j.media.2016.08.004
- Yoo, Y., Tang, L. Y., Brosch, T., Li, D. K., Kolind, S., Vavasour, I., . . . Tam, R. C. J. N. C. (2018). Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *17*, 169-178.
- Yoon, H. J., Sohn, I., Cho, J. H., Lee, H. Y., Kim, J.-H., Choi, Y.-L., . . . Kim, J. J. M. (2015). Decoding tumor phenotypes for ALK, ROS1, and RET fusions in lung adenocarcinoma using a radiomics approach. *94*(41).
- Zech, J., Pain, M., Titano, J., Badgeley, M., Schefflein, J., Su, A., . . . Oermann, E. K. J. R. (2018). Natural language-based machine learning models for the annotation of clinical radiology reports. *287*(2), 570-580.
- Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, *14*(1), 35-62. doi:[https://doi.org/10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7)
- Zhang, Y., Guo, S.-L., Han, L.-N., & Li, T.-L. (2016). Application and Exploration of Big Data Mining in Clinical Medicine. *129*(6), 731-738. doi:10.4103/0366-6999.178019

Publikationsverzeichnis

Vortrag auf dem **98. Röntgenkongress in Leipzig**, Deutschland vom 24.- 27. Mai 2017:

*Mohammad Q, Limburg R, Krombach G, Obert M. **Künstliches neuronales Netz zur automatischen Markierung von Lungenflügeln in CT-Datensätzen.** 98. Deutscher Röntgenkongress, Fortschr Röntgenstr 2017; 189(S 01): S1-S124 DOI: 10.1055/s-0037-1600181*

Vortrag an der **Harvard Medical School, Boston, USA** (Massachusetts Eye and Ear), *Auditory Brainstem Implants and Optical Stimulation Lab*, am 27.02.2018

*Mohammad Q, Limburg R, Krombach G, Obert M. **Deep Learning: automated classification of CT-chest-data in healthy and fibrosis.***

Ehrenwörtliche Erklärung

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.“

Ort, Datum

Unterschrift

Danksagung

Ein besonderer Dank gilt Frau Prof. Dr. med. Gabriele A. Krombach für ihre konstruktive Kritik und ihrer kritischen Auseinandersetzung mit dem Thema. Vielen Dank für die sehr differenzierten und professionellen Anmerkungen zum Vortrag auf dem Röntgenkongress.

Vom ganzen Herzen möchte ich Herrn Dr. Martin Obert danken. Ohne ihn wäre die Arbeit nicht zu Stande gekommen. Er stand mir jederzeit, sowohl fachlich, aber auch persönlich mit Rat und Tat zur Seite. Seine unendliche Geduld, liebevolles Verständnis, sehr differenzierte Anmerkungen, aber vor allem der moralische Beistand und menschliche Halt waren Hauptsäulen dieser Arbeit. Er hat mir die notwendige Kraft und den notwendigen Mut gegeben, diese Arbeit zu beginnen und zu vollenden. Dr. Martin Obert hat in mir nachhaltiges, über diese Arbeit hinaus reichendes Interesse für die künstliche Intelligenz geweckt. Herzlichsten Dank dafür.

Tief verbunden bin ich zudem meiner Familie, die mir jederzeit unterstützend beistand, mir Kraft gab und mich ermutigte an meinen Zielen festzuhalten.

