

Ostbayerische Technische Hochschule Amberg-Weiden
Fakultät Elektrotechnik, Medien und Informatik

Studiengang IT und Automation

Schwerpunkt Data Analytics und Künstliche Intelligenz

Masterarbeit

von

Jan Raber

**Entwicklung eines Deep Learning-basierten
Assistenzsystems zur Klassifikation von Pneumonien
anhand von Röntgenbildern**

A Deep Learning Based Assistance System for the
Classification of Pneumonia Using X-ray Images

Ostbayerische Technische Hochschule Amberg-Weiden
Fakultät Elektrotechnik, Medien und Informatik

Studiengang IT und Automation

Schwerpunkt Data Analytics und Künstliche Intelligenz

Masterarbeit

von

Jan Raber

Entwicklung eines Deep Learning-basierten Assistenzsystems zur Klassifikation von Pneumonien anhand von Röntgenbildern

A Deep Learning Based Assistance System for the Classification of Pneumonia Using X-ray Images

Bearbeitungszeitraum: von 1. Dezember 2020
bis 31. Mai 2021

1. Prüfer: Prof. Dr. Fabian Brunner

2. Prüfer: Dr. Guido-Arndt Söldner

Bestätigung gemäß § 12 APO

Name und Vorname
der Studentin/des Studenten: **Raber, Jan**

Studiengang: **IT und Automation**

Ich bestätige, dass ich die Masterarbeit mit dem Titel:

**Entwicklung eines Deep Learning-basierten Assistenzsystems zur Klassifikation
von Pneumonien anhand von Röntgenbildern**

selbstständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt, keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Datum: 31. Mai 2021

Unterschrift:

Masterarbeit Zusammenfassung

Studentin/Student (Name, Vorname): **Raber, Jan**
Studiengang: IT und Automation
Aufgabensteller, Professor: Prof. Dr. Fabian Brunner
Durchgeführt in (Firma/Behörde/Hochschule): Capgemini Deutschland GmbH
Betreuer in Firma/Behörde: Dr. Eldar Sultanow
Ausgabedatum: 1. Dezember 2020
Abgabedatum: 31. Mai 2021

Titel:

Entwicklung eines Deep Learning-basierten Assistenzsystems zur Klassifikation von Pneumonien anhand von Röntgenbildern

Zusammenfassung:

Eine Pneumonie, auch Lungenentzündung genannt, ist eine akute oder chronische Entzündung des Lungengewebes und gehört weltweit zu den häufigsten tödlich verlaufenden Infektionserkrankungen. Bei der Diagnostizierung der Krankheit werden in der Regel Röntgenbilder des menschlichen Oberkörpers analysiert. Daher wurde im Rahmen der vorliegenden Arbeit ein Deep Learning-basiertes Vorhersagemodell zur Erkennung von Pneumonien in Röntgenbildern entwickelt. Für das Training wurden 63 500 geeignete Röntgenbilder aus ursprünglich etwa einer Millionen Bilder extrahiert. Um ein möglichst unvoreingenommenes und robustes Modell zu erhalten, wurde die Analyse und Stratifikation der Trainingsdaten besonders fokussiert. Zur Demonstration der Praxisrelevanz wurde eine Webanwendung als Benutzeroberfläche entwickelt, sowie eine Bereitstellung des Systems in der Cloud durchgeführt. Ziel der Arbeit ist es, ein prototypisches Assistenzsystem zu entwickeln, welches das Potential künstlicher Intelligenz für die Analyse von Röntgenbildern veranschaulicht.

Schlüsselwörter: Deep Learning, Pneumonie, Assistenzsystem, Cloud Computing, Röntgen

Inhaltsverzeichnis

1 Einleitung	1
1.1 Ausgangssituation	2
1.2 Motivation und Zielsetzung	3
1.3 Aufbau der Arbeit	4
2 Einführung in bildgebende Verfahren	5
2.1 Entdeckung und Geschichte	6
2.2 Projektionsröntgen	7
2.2.1 Erzeugung der Röntgenstrahlung	7
2.2.2 Durchdringung von Materie	8
2.2.3 Techniken zur Bildaufnahme	8
2.2.4 Verbreitung	9
2.2.5 Kosten	10
2.3 Computertomographie	10
2.3.1 Funktionsweise	10
2.3.2 Verbreitung	11
2.3.3 Kosten	12
3 Medizinische Grundlagen	13
3.1 Diagnoseprozess	13
3.2 Assistenzsysteme	15
3.3 Diagnose von Pneumonien	16
3.3.1 Die Thoraxaufnahme	16
3.3.2 Röntgenzeichen einer Pneumonie	17
3.3.3 Computertomographie zur Pneumoniediagnose	20
3.4 Perspektivische Eigenschaften von Röntgenbildern	20
3.4.1 Lage- und Richtungsbezeichnungen	21
3.4.2 Bildperspektiven	22
4 Auswahl und Modellierung der Daten	26
4.1 Einfluss der Eigenschaften	26
4.2 Analyse und Bereinigung der Datensätze	27
4.2.1 Open-i	29
4.2.2 ChestX-ray14	30
4.2.3 Pädiatrie Datensatz (Kermany)	32
4.2.4 CheXpert	33
4.2.5 MIMIC-CXR	37
4.2.6 PadChest	40

4.3	Auswahl der Daten	44
4.3.1	Balancierung der Klassen	44
4.3.2	Stratifikation	45
4.3.3	Kombination der Datensätze	49
5	Vorhersagemodell	51
5.1	Theoretische Grundlagen	51
5.1.1	Convolutional Neural Networks	51
5.1.2	Transfer Learning	52
5.1.3	Verwendete Bibliotheken	53
5.2	Datenvorverarbeitung	54
5.2.1	Entfernen von Seitenrändern	55
5.2.2	Quadratisches Seitenverhältnis	55
5.2.3	Skalierung der Bilder	56
5.3	Modelltraining	57
5.3.1	Verifikation des ursprünglichen Modells	57
5.3.2	Einteilung der Daten	58
5.3.3	Augmentation	59
5.3.4	Einfluss während des Trainings	59
5.3.5	Modellarchitektur	60
5.4	Bewertung des Modells	61
5.4.1	Vorhersagegenauigkeit	61
5.4.2	Generalisierbarkeit	62
5.4.3	Zielvariable	64
5.4.4	Multi-Label-Klassifizierung	65
6	Grafische Benutzeroberfläche	66
6.1	Ziel und Funktionsumfang	66
6.1.1	Zielgruppe	67
6.1.2	Selbsterklärendes System	67
6.1.3	Analyse von Röntgenbildern	68
6.1.4	Optimierung für mobile Endgeräte	68
6.1.5	Demonstration der Erweiterbarkeit	68
6.2	Verwendete Technologien	69
6.2.1	Angular Framework	69
6.2.2	Externe Bibliotheken	70
6.3	Aufbau der Anwendung	70
6.3.1	Architektur	70
6.3.2	Startseite und Tutorial	74
6.3.3	Analyse von Röntgenbildern	75
6.3.4	Patientenübersicht	78
6.3.5	Mobile Ansicht	79
6.4	Vergleich mit anderen Ansätzen	80
6.4.1	Native Anwendungen	80
6.4.2	Lokale Klassifikation	81

7 Bereitstellung des Systems	82
7.1 Verwendete Cloud-Computing-Dienste	82
7.1.1 Simple Storage Service	82
7.1.2 CloudFront	83
7.1.3 Certificate Manager	83
7.1.4 Elastic Load Balancing	83
7.1.5 Elastic Container Registry	84
7.1.6 Elastic Container Service	84
7.2 Umsetzung	85
7.2.1 Backend	85
7.2.2 Frontend	87
7.2.3 Architektur	89
7.3 Datenschutz	92
8 Zusammenfassung und Ausblick	93
8.1 Zusammenfassung der Arbeit	93
8.2 Limitationen	94
8.3 Ausblick	94
Literaturverzeichnis	96
Abbildungsverzeichnis	103
Tabellenverzeichnis	106
Quellcodeverzeichnis	107

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich bei der Anfertigung der Masterarbeit unterstützt haben.

Ich bedanke mich in erster Linie bei Herrn Dr. Guido-Arndt Söldner und Herrn Prof. Dr. Fabian Brunner für die Betreuung und Korrektur der Arbeit seitens der Ostbayerischen Technischen Hochschule Amberg-Weiden.

Des Weiteren möchte ich mich bei der Capgemini Deutschland GmbH bedanken, die es mir ermöglichte die Arbeit im professionellen Umfeld der Niederlassung Nürnberg durchzuführen.

Abkürzungsverzeichnis

A | B | C | D | E | J | K | L | M | N | O | P | R | S | T | W | Y

A

ALB	Application Load Balancer
AP	Antero-Posterior
API	Application Programming Interface
AUROC	Area Under the Receiver Operating Characteristic Curve
AWS	Amazon Web Services

B

BIMCV	Banco digital de Imagen Medica de la Comunidad Valenciana
--------------	---

C

CAD	Computer Aided Diagnosis
CCD	Charge-Coupled Device
CDN	Content Delivery Network
CNN	Convolutional Neural Network
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
CT	Computertomographie
CXR14	ChestX-ray14

D

DI	Dependency Injection
DICOM	Digital Imaging and Communications in Medicine

DSGVO Datenschutz-Grundverordnung

E

EC2 Elastic Compute Cloud

ECR Elastic Container Registry

ECS Elastic Container Service

ELB Elastic Load Balancing

J

JPEG Joint Photographic Experts Group

JSON JavaScript Object Notation

K

KI Künstliche Intelligenz

L

L Lateral

M

MeSH Medical Subject Headings

MIT Massachusetts Institute of Technology

ML Machine Learning

MRT Magnetresonanztomographie

N

NIH National Institutes of Health

NLM National Library of Medicine

NLP Natural Language Processing

O

OCT Optische Kohärenztomographie

OOD	out-of-distribution
OpenCV	Open Source Computer Vision Library

P

PA	Postero-Anterior
PET	Positronen Emissions Tomographie
PHP	PHP: Hypertext Preprocessor
PNG	Portable Network Graphics

R

RadLex	Radiology Lexicon
REST	Representational State Transfer
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic

S

S3	Simple Storage Service
SDK	Software Development Kit
SPECT	Single Photon Emission Computed Tomography

T

TLS	Transport Layer Security
------------	--------------------------

W

WHO	Weltgesundheitsorganisation
------------	-----------------------------

Y

YAML	YAML Ain't Markup Language
-------------	----------------------------

Kapitel 1

Einleitung

Eine Pneumonie, auch Lungenentzündung genannt, ist eine akute oder chronische Entzündung des Lungengewebes, welche meist durch eine Infektion mit Viren, Bakterien oder Pilzen verursacht wird. Weltweit gehört die ambulant erworbene Pneumonie zu den häufigsten tödlich verlaufenden Infektionserkrankungen. Im Gegensatz zu vielen anderen Infektionskrankheiten, sind auch die Bevölkerungen entwickelter Nationen massiv von der Krankheit betroffen. Basierend auf Statistiken der Bundesgeschäftsstelle für Qualitätssicherung wird die jährliche Gesamthäufigkeit von Pneumonien in Deutschland auf etwa 400 000 bis 600 000 geschätzt [25]. Jedes Jahr sterben laut Todesursachenstatistik des Bundesamts für Statistik knapp 20 000 Menschen an einer Lungenentzündung, wobei die Letalität mit steigendem Alter drastisch zunimmt [77].

Durch die schlechtere medizinische Versorgung sind schwere und tödliche Krankheitsverläufe in Entwicklungsländern jedoch wesentlich häufiger. Diese Problematik lässt sich an einer globalen Studie der Weltgesundheitsorganisation (WHO) erkennen, welche die Todesursachen von Kindern bis fünf Jahren untersuchte. Die Schätzungen ergaben, dass etwa 46 % der Todesfälle, die auf Pneumonien zurückzuführen waren, in Afrika verzeichnet wurden [9].

Bei der Diagnostizierung von Pneumonien und weiteren Lungenkrankheiten werden in der Regel Thoraxröntgenbilder eingesetzt. Situationsabhängig werden diese durch aufwändiger Verfahren wie Computertomographie (CT) oder Labortests ergänzt. Die öffentliche Verfügbarkeit medizinischer Bilddaten in Form von Datensätzen hat in den letzten Jahren stark zugenommen. Diese Entwicklung ermöglicht es, den rasanten Fortschritt im Bereich des maschinellen Sehens (englisch: *Computer Vision*) auch auf medizinische Aufgaben anzuwenden. Unter maschinellem Sehen versteht man dabei die computerbasierte Interpretation von Bilddaten, häufig unter Verwendung von Deep Learning.

Um die Diagnostizierung von Lungenentzündungen zu erleichtern, soll daher im Rahmen dieser Arbeit ein Assistenzsystem entwickelt werden, welches die Präsenz von Pneumonien in Thoraxröntgenbildern feststellen soll. Im nachfolgenden Abschnitt wird die Ausgangssituation erläutert, um eine Abgrenzung der geleisteten Arbeit zu ermöglichen. Anschließend wird die Motivation und Zielsetzung vorgestellt, um letztlich die Gliederung der Arbeit darzulegen.

1.1 Ausgangssituation

Im Vorfeld dieser Arbeit wurde eine Vorgehensweise zur Optimierung der Vorhersagegenauigkeit neuronaler Netze bei Thoraxröntgenbildern im Rahmen einer Studienarbeit untersucht. Das dabei entstandene Modell wurde nicht im Rahmen dieser Arbeit entwickelt und wird im Folgenden beschrieben, um ein Verständnis für die Ausgangssituation der Arbeit zu ermöglichen.

Der Grundgedanke bestand darin, einem neuronalen Netz zusätzlich zum eigentlichen Röntgenbild die Bildbereiche erneut zur Verfügung zu stellen, die für die Erkennung einer Lungenentzündung ausschlaggebend sind. Da Pneumonien in Röntgenbildern anhand des Erscheinungsbilds der Lunge diagnostiziert werden, wurde ein neuronales Netz zur Segmentierung der Lungenregionen trainiert. Für diesen Zweck wurde die U-Net Architektur verwendet, welche eine Convolutional Neural Network (CNN)-Architektur darstellt, die speziell für die Segmentierung medizinischer Bilddaten entwickelt wurde [70]. Als Trainingsdaten wurden zwei Datensätze mit manuell segmentierten Lungen verwendet, die von der National Library of Medicine (NLM) veröffentlicht wurden [30]. Die Daten sind öffentlich verfügbar und können über das in Abschnitt 4.2.1 beschriebene Open-i System heruntergeladen werden. Das beispielhafte Ergebnis einer Lungensegmentierung ist in Abbildung 1.1 dargestellt. Aus dem segmentierten Röntgenbild wird durch Binarisierung eine Maske erstellt, welche anschließend auf das ursprüngliche Röntgenbild angewendet wird, um die Lungenregionen auszuschneiden.

Das ursprüngliche Röntgenbild und die ausgeschnittenen Lungenregionen wurden auf zwei unterschiedlichen Kanälen in ein simples CNN eingespeist. Die beschriebene Architektur wurde für die Unterscheidung zwischen gesunden Patienten und Patienten mit Pneumonie trainiert. Für den Trainingsprozess wurde der in Abschnitt 4.2.3 beschriebene Datensatz von Kermany u. a. verwendet. Das Vorhersagemodell erreichte dabei vielversprechende Ergebnisse, welche aufgrund des sehr homogenen und kleinen Datensatzes jedoch nur begrenzt aussagekräftig waren.

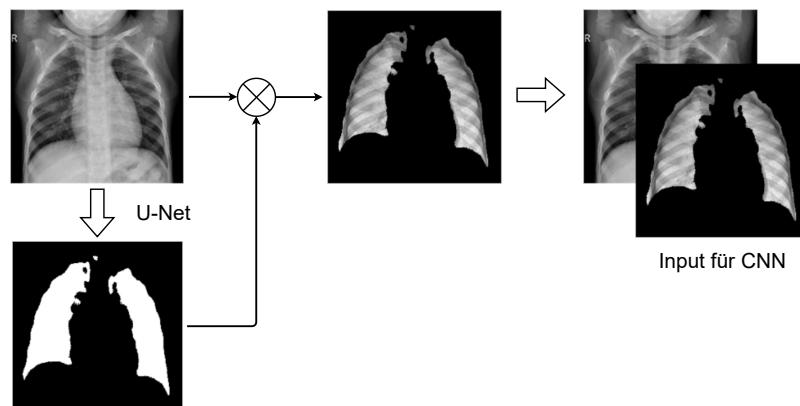


Abbildung 1.1: Ursprüngliche Modellarchitektur zur Klassifikation von Pneumonien (Quelle des Röntgenbilds: [36])

1.2 Motivation und Zielsetzung

Wie zu Beginn des Kapitels bereits angesprochen wurde, sind Lungenentzündungen weltweit eine der häufigsten Todesursachen, obwohl sie durch herkömmliche Röntgenaufnahmen diagnostiziert werden können. Fehldiagnosen bei der Begutachtung der Röntgenbilder sind dabei häufig auf unzureichende Expertise oder Zeitmangel zurückzuführen. Besonders während der COVID-19 Pandemie zeigte sich, dass es dem medizinischen Fachpersonal durch die Überlastung der Gesundheitssysteme nicht immer möglich war, eine ausreichende Diagnostizierung und Behandlung aller Patienten sicherzustellen.

Ziel dieser Arbeit ist es daher ein prototypisches Assistenzsystem zu entwickeln, das Mediziner in schwierigen Situationen unterstützt, indem es innerhalb von Sekunden eine zweite Meinung zur Verfügung stellt. In der Fachsprache spricht man dabei oft von Computer Aided Diagnosis (CAD)-Systemen. Wie der Begriff bereits vermuten lässt, soll die Rolle des Systems ausschließlich unterstützend und assistierend sein. Bei unsicheren Entscheidungen könnte das Modell den behandelnden Arzt beispielsweise in seiner Verdachtsdiagnose bekräftigen oder dafür sorgen, dass der Zustand des Patienten mit weiteren Untersuchungen genauer bestimmt werden muss.

Bei der Analyse des aktuellen Forschungsstandes zeigte sich, dass bereits eine Vielzahl unterschiedlicher Deep Learning-Ansätze erfolgreich auf medizinische Klassifizierungsaufgaben angewendet werden konnte. Besonders auffällig ist hierbei, dass die Vorhersagemodelle meist nur auf einem Datensatz trainiert wurden, was in einer geringen Diversität der Daten resultiert. Im wissenschaftlichen Bereich ist dies eine gängige Vorgehensweise, die eine leichtere Einordnung der Ergebnisse und bessere Vergleichbarkeit der unterschiedlichen Arbeiten ermöglicht. Auf die praktische Einsatzbarkeit der Modelle wirkt sich diese Methodik jedoch kontraproduktiv aus, da die Datenlage in realen Anwendungsfällen deutlich heterogener ist.

Die gewonnenen Erkenntnisse sollten daher auf vielfältigere Daten angewendet werden, um die Generalisierbarkeit und Robustheit der Vorhersagemodelle zu optimieren. Hierfür sollen mehrere öffentlich verfügbare Datensätze kombiniert werden, um anschließend eine Verifikation des aktuellen Stands der Technik auf diesen Daten durchzuführen. Somit kann das Modell mit Röntgenbildern trainiert werden, die in möglichst unterschiedlichen Krankenhäusern und Rahmenbedingungen entstanden sind. Da medizinische Datensätze häufig einen nicht zu vernachlässigenden Bias vorweisen, sollte bei der Auswahl der Daten besonders auf die Gegebenheiten der Metainformationen und demografischen Eigenschaften geachtet werden.

Weiterhin soll der Prototyp den Diskurs zwischen Medizin und Informatik anregen. Ein praxisnahes Beispielsystem kann beiden Seiten helfen, technische Möglichkeiten und medizinische Anforderungen verständlich zu kommunizieren. Für das Assistenzsystem soll eine intuitiv bedienbare Benutzeroberfläche entwickelt werden, welche die Analyse von Röntgenbildern ermöglicht. Um flexibel einsetzbar zu sein, sollte die Anwendung möglichst unabhängig vom Betriebssystem oder der darunterliegenden

Hardware sein. Zusätzlich soll ein Konzept erarbeitet und umgesetzt werden, welches die Bereitstellung des Systems ermöglicht.

Das prototypische CAD-System kann somit einen wichtigen Beitrag zur Förderung der praktischen Einsetzbarkeit von Deep Learning im medizinischen Sektor leisten. Die Visualisierung und Interaktion kann es zudem ermöglichen, dass dieser bedeutsame Forschungsbereich einem breiteren Publikum zugänglich wird, da für die Verwendung und das Verständnis des Systems keine Vorkenntnisse erforderlich sind.

1.3 Aufbau der Arbeit

Im zweiten Kapitel dieser Arbeit werden grundlegende Kenntnisse zu bildgebenden Verfahren erarbeitet, um ein Verständnis für die Erzeugung von Thoraxröntgenbildern zu ermöglichen. Anschließend wird der medizinische Diagnoseprozess in Kapitel 3 vorgestellt. Die Diagnostizierung von Pneumonien und die Einbringung von CAD-Systemen in den Diagnoseprozess werden dabei besonders fokussiert. Kapitel 4 beschreibt die Extraktion und Modellierung relevanter Röntgenbilder aus den sechs verwendeten Datensätzen und ist das wichtigste Kapitel dieser Arbeit, da die Auswahl der Bilder ausschlaggebend für die Generalisierbarkeit und Praxisrelevanz des Vorhersagemodells ist. Die Anwendung der aktuell besten Vorgehensweisen zur Klassifizierung von Thoraxröntgenbildern wird im darauffolgenden Kapitel erläutert. In den Kapiteln 6 und 7 wird die Architektur und Implementierung des Assistenzsystems beschrieben, sowie dessen Bereitstellung mittels Cloud-Computing vorgestellt. Abschließend werden die Ergebnisse und Limitationen der Arbeit zusammengefasst und ein Ausblick über die mögliche Weiterentwicklung der Vorhersagemodele und des Assistenzsystems gegeben.

Kapitel 2

Einführung in bildgebende Verfahren

Bei vielen medizinischen Fragestellungen ist die Verwendung bildgebender Verfahren unabdingbar. Befindet sich Flüssigkeit in einem der Lungenflügel des Patienten? Hat sich eine Person nach einem Sturz eine Fraktur zugezogen? Leidet der Patient an einer Krebserkrankung? Diese und viele weitere Fragen können durch Bilder vom Inneren des menschlichen Körpers häufig präzise beantwortet werden. Seit der Entdeckung der Röntgenstrahlung, die als Grundstein der bildgebenden Verfahren gilt, wurden viele unterschiedliche Technologien entwickelt. Die nachfolgende Auflistung bietet eine kurze Übersicht der gängigsten Verfahren, gruppiert nach der physikalischen Grundlage der Bilderzeugung.

- Röntgenstrahlung (z.B. Projektionsröntgen, CT, Mammographie)
- Kernspinresonanz (z.B. Magnetresonanztomographie (MRT))
- Nuklearmedizin (z.B. Szintigraphie, Positronen Emissions Tomographie (PET), Single Photon Emission Computed Tomography (SPECT))
- Impedanz (z.B. Impedanztomographie)
- Ultraschall (z.B. Sonographie)
- Infrarotstrahlung (z.B. Thermographie)
- Sichtbares Licht (z.B. Endoskopie, Optische Kohärenztomographie (OCT))

Welche Vorgehensweise zum Erlangen der benötigten Informationen die beste ist, hängt hauptsächlich von der jeweiligen medizinischen Fragestellung ab. Dieses Kapitel soll eine Einführung in bildgebende Verfahren geben, da dieses Wissen zum Verständnis des vorliegenden Klassifikationsproblems benötigt wird. Da die vorliegende Arbeit die maschinelle Diagnose von Pneumonien behandelt, beschränkt sich dieses Kapitel ausschließlich auf Technologien, die zur Pneumoniediagnose eingesetzt werden. Die Röntgendiagnostik wird hierbei besonders fokussiert, da diese für die Arbeit von besonderer Bedeutung ist.

2.1 Entdeckung und Geschichte

Die Röntgenstrahlung wurde 1895 von Wilhelm Conrad Röntgen entdeckt, als er fluoreszenzfähige Gegenstände beobachtete, die nahe einer abgedeckten Crookes-Röhre zu leuchten begannen. Nach seiner außergewöhnlichen Beobachtung begann Röntgen deren Ursache wissenschaftlich zu untersuchen. Schnell wurde klar, dass es sich hierbei um eine neuartige Strahlung handelt, welche beim Anlegen hoher Spannungen von der Röhre ausgesendet wird. Während der Untersuchungen stellte er fest, dass Röntgenstrahlung Materie abhängig vom Material unterschiedlich gut durchdringen kann [71]. Dass die Strahlung je nach Material unterschiedlich stark geschwächt wird, ist der Grund, weshalb es überhaupt möglich ist, Lungenentzündungen anhand von Röntgenbildern zu diagnostizieren. Röntgen erkannte bereits nach kürzester Zeit das Potential seiner Entdeckung für Einsatzmöglichkeiten im medizinischen Bereich. So fertigte er nur wenige Wochen nach Entdeckung der neuen Strahlungsart die ersten Röntgenbilder von menschlichen Händen an. Nach einem Vortrag an der Physikalisch-Medizinischen Gesellschaft in Würzburg entstand die weltbekannte Röntgenaufnahme der Hand von Prof. Rudolf Albert von Koelliker (Abbildung 2.1). Durch die einfach verständliche Visualisierung der Technologie wurde das Ausmaß der Entdeckung auch schnell außerhalb wissenschaftlicher Kreise wahrgenommen. Dass nun beispielsweise Frakturen und innere Verletzungen nichtinvasiv erkannt und untersucht werden konnten, war eine Revolution in der medizinischen Diagnostik. Röntgen erhielt für die Entdeckung der nach ihm benannten Röntgenstrahlen 1901 den ersten Nobelpreis für Physik [54]. Selbstverständlich wurden seit der Entdeckung sämtliche Bestandteile der Röntgengeräte und -diagnostik fortlaufend weiterentwickelt und verbessert. Trotz der vielen Innovationen basieren moderne Röntgengeräte noch heute auf dem zugrundeliegenden Prinzip der Durchleuchtung des menschlichen Körpers durch Röntgenstrahlen und ihrer materialabhängigen Schwächung.



Abbildung 2.1: Historische Aufnahme der Hand von Prof. Rudolf Albert von Koelliker aus 1896, aufgenommen im Physikalischen Institut der Universität Würzburg. (Quelle: [71])

2.2 Projektionsröntgen

Das Projektionsröntgen ist nach wie vor eines der wichtigsten bildgebenden Verfahren in der modernen Medizin. Es liefert gute Antworten auf viele diagnostische Fragen und ist dabei schnell, flexibel und preiswert. Zu den klassischen Anwendungsgebieten des Röntgen gehören die Untersuchung des knöchernen Skeletts, sowie der inneren Organe und Gefäße.

2.2.1 Erzeugung der Röntgenstrahlung

Grundsätzlich werden in einer Röntgenröhre Elektronen freigesetzt und mit einer hohen Spannung (etwa im Bereich 80kV bis 150kV) von der Kathode hin zur Anode beschleunigt. Die in der Anode beim Auftreffen der Elektronen freigesetzten Prozesse erzeugen die Röntgenstrahlung. Um die physikalischen Prozesse des Röntgens im Detail zu verstehen, bietet sich „Bildgebende Verfahren in der Medizin“ von Olaf Dössel an, der sich mit seinem Buch vor allem an Physiker, Ingenieure und Informatiker wendet [17, S. 5-83]. Um ein qualitativ hochwertiges Röntgenbild zu erzeugen, muss die Energie der Röntgenröhre abhängig vom zu untersuchenden Gewebe verändert und die Fokussierung der Körperregion richtig eingestellt werden. Somit ist die Kalibrierung des Geräts abhängig von der Art der diagnostischen Fragestellung.

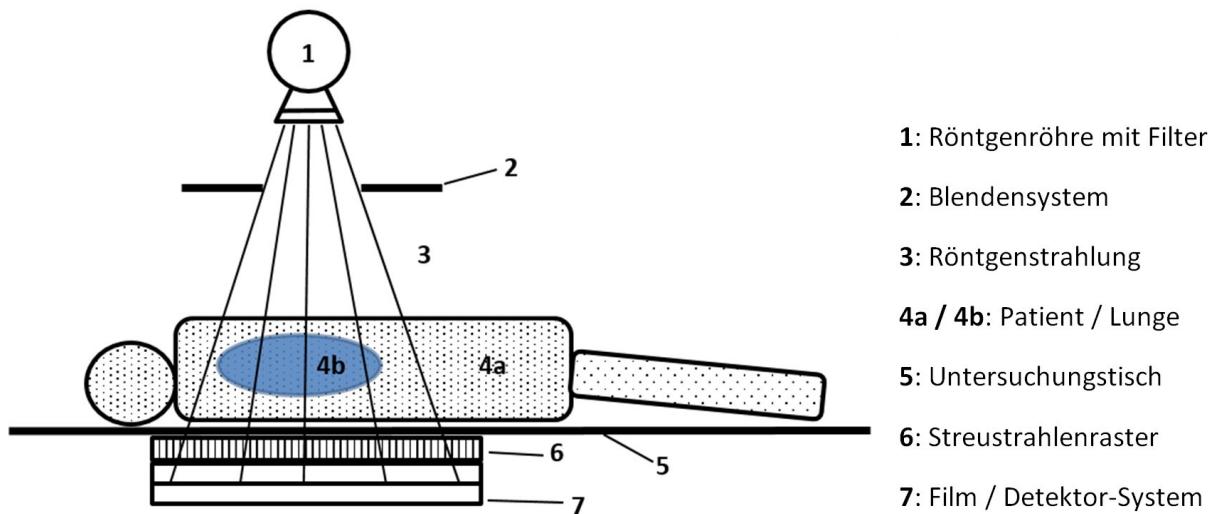


Abbildung 2.2: Die Komponenten des Röntgensystems bei der Anfertigung einer Thoraxaufnahme
(Quelle: [79])

Der Anteil der Röntgenstrahlung mit langen Wellenlängen wird auch als weiche Röntgenstrahlung bezeichnet. Diese wird vom menschlichen Körper vollständig absorbiert (siehe Abschnitt 2.2.2) und trägt daher nicht zur Bildgebung bei. Da Röntgenstrahlung in zu hoher Dosierung gesundheitliche Schäden verursachen kann, ist es besonders wichtig, die Strahlenbelastung des Patienten möglichst gering zu halten. Zur Reduzierung der Strahlenbelastung werden daher Aluminiumfilter eingesetzt, welche die nicht benötigten Anteile der Röntgenstrahlung absorbieren (siehe Abbildung 2.2, (1)).

Um nur die Körperregionen der Strahlung auszusetzen, die untersucht werden sollen, kommt zudem ein Blendensystem zum Einsatz (siehe Abbildung 2.2, (2)). Zusätzlich zur Blende irrelevante Körperteile mit Bleischürzen abgedeckt, um die Strahlenbelastung des Patienten zu verringern [vgl. 17, S. 22].

2.2.2 Durchdringung von Materie

Beim Auftreffen der Röntgenstrahlen auf Materie kommt es zu verschiedenen Wechselwirkungen. Im Rahmen dieser Arbeit werden diese physikalischen Effekte vereinfacht beschrieben und auf die Schwächung, Absorption und Streuung der Strahlen begrenzt. Beim Durchdringen eines Materials durch einen Röntgenstrahl wird dieser geschwächt. Als materialabhängige Konstante für die Stärke der Schwächung dient der sogenannte Schwächungskoeffizient μ . Die Schwächung der Strahlen ist hierbei weitgehend proportional zur Dichte des Materials. Für ein dickes Material mit homogener Zusammensetzung gilt das allgemeine Schwächungsgesetz

$$I = I_0 \cdot e^{-\mu \cdot d} \quad (2.1)$$

mit:

I = Intensität der Strahlung nach dem Materialdurchgang

I_0 = Intensität vor dem Materialdurchgang

μ = Schwächungskoeffizient

d = Dicke des Materials

Gibt der Röntgenstrahl seine gesamte Energie an das Material, kommt es nicht nur zu einer Schwächung, sondern sogar zu einer Absorption. Röntgenstrahlen, die von der Materie absorbiert wurden, treffen nicht auf den Detektor und können somit nicht mehr zur Bildgebung beitragen. Ein strahlendurchlässiges Organ, etwa die Lunge, hält kaum Strahlen zurück. Folglich verlässt ein Großteil dieser Strahlen den Patienten, was zu einer Schwarzfärbung führt. Strahlenundurchlässigeres Gewebe, zum Beispiel ein Knochen, schwächt die Strahlen stärker ab, so dass auf dem Röntgenbild dort helle Strukturen zu sehen sind. Abgesehen von der Schwächung und Absorption werden Röntgenstrahlen beim Durchdringen eines Materials gestreut. Bei der Darstellung des Körperinneren werden die Strahlen abhängig vom Gewebe unterschiedlich stark absorbiert und gestreut.

2.2.3 Techniken zur Bildaufnahme

Nachdem die Strahlen den Körper durchdrungen haben, muss noch das eigentliche Röntgenbild angefertigt werden. Die in Abschnitt 2.2.2 beschriebene Streustrahlung ist für die Aufnahme des Röntgenbilds ein Störfaktor. Für die Bildgebung sollen daher möglichst nur Strahlen berücksichtigt werden, die den Körper ohne Winkeländerung

durchdrungen und auf der gegenüberliegenden Seite geschwächt wieder ausgetreten sind. Um die Bildqualität zu verbessern, werden daher sogenannte Streustrahlenraster eingesetzt (siehe Abbildung 2.2, (6)). Das Raster besteht abwechselnd aus dünnen Streifen eines stark absorbierenden Materials (z.B. Bleifolie) und dickeren Streifen eines sehr durchlässigen Materials. Die Streifen sind hierbei parallel zur Strahlung ausgerichtet, die von der Röntgenröhre ausgesendet wird. Der Großteil der Streustrahlung ist nicht parallel zu den Streifen ausgerichtet, weshalb sie vom Streustrahlenraster absorbiert werden. Die unerwünschte Strahlung wird somit minimiert, bevor sie auf den Film beziehungsweise den Detektor auftreffen kann.

Nachdem die Röntgenstrahlung den Tisch und das Streustrahlenraster passiert hat, trifft sie auf den Detektor, wo das eigentliche Röntgenbild erzeugt wird. Bei der analogen Bildaufnahme wird als Detektor ein Röntgenfilm verwendet. Ebenso wie herkömmliche fotografische Filme, ist auch der Röntgenfilm ein sogenannter Negativfilm. Das heißt, je mehr Röntgenstrahlung auf den Film trifft, desto dunkler färbt er sich. In der Anfangszeit wurden noch keine Negativfilme verwendet, weshalb die Knochen in Abbildung 2.1 dunkel dargestellt sind, während sie in heutigen Röntgenbildern hell in Erscheinung treten.

Der Entwicklungsprozess der Bilder wird zunehmend digitalisiert. Um ältere Röntgengeräte trotz fortschreitender Digitalisierung verwenden zu können, wird häufig ein hybrider Ansatz gewählt. Dabei wird im ersten Schritt ein analoges Röntgenbild auf einem Röntgenfilm aufgenommen, um dieses anschließend mit einem speziellen Scanner in eine Datei zu konvertieren. Bei modernen Röntgengeräten geschieht die Aufnahme des Bilds durch einen digitalen Röntgen-Bildaufnehmer. Die verwendete Technologie ähnelt dabei der Charge-Coupled Device (CCD)-Sensorik einer herkömmlichen Digitalkamera. Da die Röntgenstrahlen selbst nicht von dem Sensor erfasst werden können, wird eine Lumineszenzschicht verwendet, welche die Strahlen in sichtbares Licht konvertiert [vgl. 17, S. 51]. Die digitale Bilderzeugung sorgt neben der weit besseren Anwenderfreundlichkeit auch für eine höhere Bildqualität und weniger Strahlenbelastung für den Patienten.

2.2.4 Verbreitung

In entwickelten Ländern ist die Röntgendiagnostik ein Standardmittel, das beispielsweise in Kliniken, bei Zahnärzten und selbst bei Tierärzten täglich eingesetzt wird. Alleine in den USA kam das Projektionsröntgen im Jahr 2018 schätzungsweise 214 Millionen Mal zum Einsatz [88]. Der Anteil an Röntgenuntersuchungen liegt damit bei 65 % innerhalb der fünf gängigsten Verfahren (Projektionsröntgen, CT, MRT, PET und Nuklearmedizin). Die WHO geht davon aus, dass circa 70-80 % der diagnostischen Problemstellungen mit einfacher medizinischer Bildgebung wie Ultraschall und Projektionsröntgen gelöst werden können [vgl. 48]. Leider zeigt der Artikel auch, dass nach Schätzungen nur etwa zwei Drittel der Weltbevölkerung Zugang zu diesen Technologien und der notwendigen Expertise hat.

Ein Beispiel für diese Knappheit lässt sich in Liberia erkennen. Liberia ist ein west-afrikanisches Land mit einer Gesamtbevölkerung von knapp 5 Millionen Einwohnern [82]. Dort gab es im Jahr 2011 im ganzen Land lediglich 56 Röntgengeräte [vgl. 47, S. 17], für deren Benutzung nur 22 Röntgenassistenten ausgebildet waren [vgl. 47, S. 30], obwohl die WHO mindestens einen Röntgenassistenten pro Gerät empfiehlt.

Noch knapper als die Geräte und die Assistenten sind die Ärzte selbst. Laut einer Studie der University of Massachusetts Medical School praktizierten in Liberia im Jahr 2015 nur 2 ausgebildete Radiologen, die zur Analyse von Röntgenbildern zertifiziert sind [vgl. 2, S. 4]. Doch selbst in wirtschaftlich stärkeren Entwicklungsländern wie Ruanda ist die Expertise in medizinischer Bildgebung noch stark eingeschränkt. Im Jahr 2015 praktizierten in Ruanda lediglich 11 Radiologen, bei einer Bevölkerung von etwa 12 Millionen Menschen [72]. Durch die fortschreitende Verbreitung der Technologie und Schulung des Personals sollten diese Diskrepanzen in Zukunft jedoch weiter abnehmen.

2.2.5 Kosten

Da eine Vielzahl unterschiedlicher Hersteller und Technologien auf dem Markt zu finden sind, fällt es schwer eine generelle Aussage über die Kosten eines Röntgengeräts zu treffen. Ein grob abgesteckter Bereich von Basismodellen, bis hin zu hochtechnologischen Geräten kann jedoch definiert werden. Für günstige Röntgengeräte beginnen die Anschaffungskosten bei mehreren zehntausend Euro. Im Gegensatz dazu kann der Preis bei besonders leistungsfähigen Geräten im obereren Segment durchaus einige hunderttausend Euro erreichen [46].

2.3 Computertomographie

Aufgrund der Praxisrelevanz (siehe Abschnitt 3.3.3) wird trotz der Fokussierung dieser Arbeit auf Röntgenbilder ein kurzer Einblick in die Computertomographie gegeben. Ziel der Technologie ist es, sogenannte Schnittbilder des Patienten anzufertigen. Beim Projektionsröntgen werden alle Ebenen des Körpers in einem Bild aufgezeichnet. Daher können Organe sich gegenseitig verdecken oder die genaue Position eines Objekts unklar sein [vgl. 17, S. 131]. Schnittbilder lösen dieses Problem, indem sie immer nur eine Ebene darstellen, die wie die Aufnahme einer dünnen Scheibe des Körpers zu sehen ist. Diese Bilder können dann entweder schrittweise vom Arzt betrachtet, oder zu einem dreidimensionalen Modell zusammengefügt werden.

2.3.1 Funktionsweise

Zur Anfertigung einer CT-Aufnahme wird eine Vielzahl aus verschiedenen Richtungen aufgenommener Röntgenbilder benötigt. Aufgrund der unterschiedlichen Winkel

ergeben sich verschiedene Perspektiven auf die gleichen Punkte im Körper. Durch die Anwendung mathematischer Transformationen und Algorithmen auf die Messsignale ist es anschließend möglich, die eingangs erwähnten Schnittbilder digital zu konstruieren.

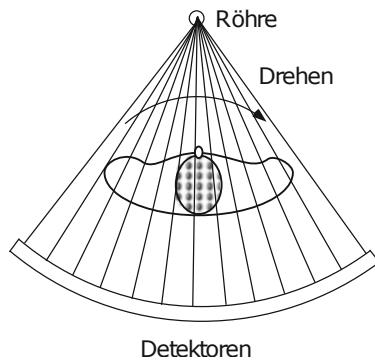


Abbildung 2.3: Prinzipieller Aufbau von weit verbreiteten CT-Scannern (Quelle: [17, S. 140])

Der schematische Aufbau weit verbreiteter CT-Scanner ist in Abbildung 2.3 zu sehen. Zur Aufnahme der Strahlung wird ein gebogenes Detektor-Array verwendet, welches hinter dem Patienten platziert ist. Die Röntgenröhre und das Detektor-Array rotieren um den Körper des Patienten und erzeugen dabei eine hohe Anzahl Schnittbilder. Diese Bilder werden anschließend entweder zur Erzeugung von Schnittbildern, oder gar zur Berechnung dreidimensionaler Modelle verwendet.

2.3.2 Verbreitung

Analog zu Abschnitt 2.2.4 wird die Verbreitung der Technologie in Industrienationen am Beispiel der USA dargelegt. Dort beträgt Anzahl der CT-Untersuchungen weniger als die Hälfte verglichen zum herkömmlichen Röntgen. Laut dem Marktforschungsunternehmen IMV wurde CT im Jahr 2019 in den USA 91 Millionen Mal eingesetzt [87]. Die geringere Nutzung der Technologie lässt sich unter anderem durch Kosten, Aufwand und diagnostische Fragestellung erklären. So ist eine CT-Untersuchung komplexer und erfordert eine aufwändigere Schulung des Personals als herkömmliche Röntgenaufnahmen. Des Weiteren können triviale und besonders eindeutige Fragestellungen in den meisten Fällen durch Ultraschall oder Projektionsröntgen diagnostiziert werden. Dennoch ist die Computertomographie eines der wichtigsten bildgebenden Verfahren und ist aus der modernen Medizin nicht mehr wegzudenken. Die zuvor beschriebene Marktforschungsstudie prognostiziert, dass die Nutzung von CT in den USA in den kommenden Jahren weiter ansteigen wird.

Ähnlich wie in anderen Bereichen der medizinischen Versorgung ist die Verbreitung in Entwicklungs- und Schwellenländern um ein Vielfaches geringer. Da in diesen Ländern bereits die Versorgung mit grundlegenden Technologien der medizinischen Bildgebung mangelhaft ist, ist die Abwesenheit von CT-Geräten die logische Konsequenz. Ebenso wie bei den Röntgengeräten scheitert die weitere Verbreitung der

Geräte in armen Regionen sowohl an den Kosten als auch an der fehlenden Expertise des medizinischen Personals. Eine Länderstudie der WHO macht diese ungleiche Verteilung der Medizingeräte deutlich [86]. Die Visualisierung der Studienergebnisse ist in Abbildung 2.4 zu sehen.

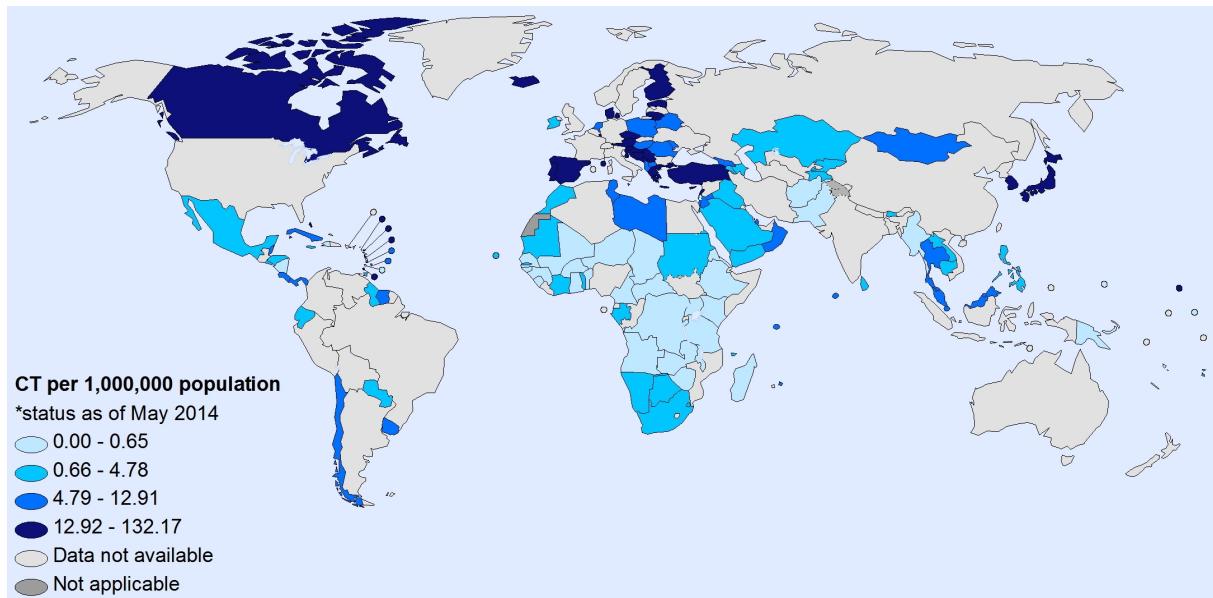


Abbildung 2.4: Länderspezifische Verbreitung von CT-Geräten, gemessen in Anzahl Geräte pro eine Millionen Einwohner (Quelle: [86])

2.3.3 Kosten

Die Anschaffungskosten für ein neues CT Gerät befinden sich im Bereich weniger hunderttausend Euro, bis hin zu einer knappen Millionen Euro [45]. Das Preisspektrum beginnt somit etwa dort, wo es für herkömmliche Röntgengeräte endet. Durch den hohen Energieverbrauch, professionelle Reinigung und weitere Wartungskosten können die laufenden Kosten eines Gerätes durchaus einige tausend Euro monatlich betragen.

Kapitel 3

Medizinische Grundlagen

Dieses Kapitel beschreibt die medizinischen Prozesse bei der Diagnostizierung von Lungenentzündungen, um ein Verständnis für die Integration des entwickelten Systems in den Diagnoseprozess zu ermöglichen. Der erste Abschnitt beschreibt allgemein welche Schritte zur Feststellung einer Krankheit vollzogen werden. Abschnitt 3.2 erläutert, wie der Diagnoseprozess durch CAD-Systeme unterstützt werden kann. Der letzte Abschnitt des Kapitels befasst sich speziell mit dem Anwendungsfall des entwickelten Systems, der Diagnose von Lungenentzündungen anhand von Röntgenbildern.

3.1 Diagnoseprozess

Als diagnostischen Prozess bezeichnet man das Sammeln von Informationen in Kombination mit klinischen Schlussfolgerungen, mit dem Ziel die Krankheit oder den gesundheitlichen Zustand eines Patienten zu bestimmen [vgl. 5, S. 32]. Der Ablauf des Prozesses ist hochgradig individuell, da er abhängig vom Patienten und dessen Symptomen ist. Aufgrund der Dynamik kann somit kein starrer Prozess definiert werden, der immer gültig ist. Es ist jedoch möglich die grundsätzlichen Abläufe und Vorgehensweisen zu beschreiben. Die in diesem Abschnitt beschriebenen Vorgänge orientieren sich an der Konzeption des Diagnoseprozesses des Komitees für diagnostische Fehler im Gesundheitswesen der gemeinnützigen Institution National Academies of Sciences, Engineering, and Medicine [5, S. 31-80].

Der Patient selbst ist in der Regel der erste, der Symptome und eine eventuell damit verbundene Krankheit feststellt. Bei besonders schweren oder länger anhaltenden Beschwerden erfolgt anschließend die Kontaktaufnahme mit dem Gesundheitssystem. Nun beginnt der eigentliche Diagnoseprozess, welcher in Abbildung 3.1 als zyklischer Vorgang dargestellt ist.

Um mögliche Ursachen der Symptome aufzudecken, müssen Daten über den Patienten und die äußereren Umstände erhoben werden. Je nach Problemstellung können viele verschiedene Methoden zur Beschaffung der Informationen angewendet werden. In der Regel wird als Erstes eine Befragung des Patienten durchgeführt, welche von diversen Untersuchungen gefolgt wird. Abhängig von der jeweiligen Situation können beispielsweise bildgebende Verfahren und Labortests zum Einsatz kommen. Die medizinische

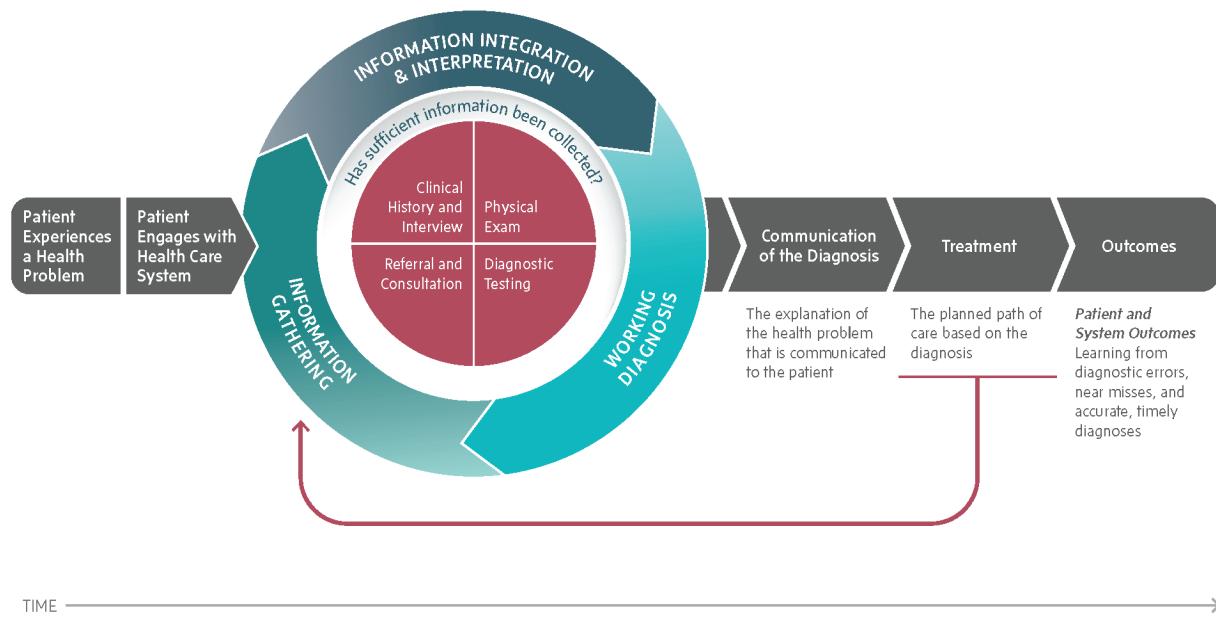


Abbildung 3.1: Diagnoseprozess vom Auftreten eines gesundheitlichen Problems bis zur Behandlung, Konzipiert vom Komitee für diagnostische Fehler im Gesundheitswesen (Quelle: [5, S. 33])

Bildgebung ist sehr flexibel und kann für die Diagnose unzähliger Erkrankungen eingesetzt werden [vgl. 5, S. 43]. Welches bildgebende Verfahren die besten Informationen liefert, hängt beispielsweise von dem zu untersuchenden Organ und dem gesundheitlichen Zustand des Patienten ab. In der medizinischen Fachsprache spricht man bei der Debatte über bildgebende Technologien auch häufig von der sogenannten Modalität. Für die gleiche Körperregion erhält man mit unterschiedlichen Technologien teils grundverschiedene Ergebnisse, die unterschiedliche Stärken und Schwächen haben. Bereits vor der Durchführung muss somit klar sein, welche Informationen sich der Arzt von der Aufnahme erhofft. Bei der Untersuchung von Gefäßen oder bestimmten Organen kann es dabei hilfreich sein, dem Patienten ein Tracer- beziehungsweise Kontrastmittel zu verabreichen. Dadurch können die hervorgehobenen Bereiche besser untersucht werden. Anschließend wird eine Befundung durchgeführt, bei der die Aufnahmeprozedur und eventuelle Auffälligkeiten in den Bildern beschrieben werden. Durch das in dieser Arbeit entwickelte CAD-System könnte sich der behandelnde Arzt während der Befundung eine zweite Meinung zu vorliegenden Röntgenbildern einholen. Abhängig von der Auslastung des Krankenhauses kann es zudem dazu kommen, dass die Befundung der Bilder nicht unmittelbar nach deren Aufnahme geschieht, da zu diesem Zeitpunkt kein Radiologe verfügbar ist. Auch hier könnte das Vorhersagemodell helfen, eine erste Einschätzung und Priorisierung der Situation zu ermöglichen, ohne Fachkenntnisse des Personals zu erfordern.

Nachdem in Gesprächen, Tests und Bilddaten Informationen über das Krankheitsbild des Patienten gesammelt wurden, müssen diese in Zusammenhang gebracht werden. Ziel ist es den gesundheitlichen Zustand des Patienten mit möglichst hoher Sicherheit zu bestimmen und die Ursache der Symptome zu identifizieren. Die aktuell wahrscheinlichste Diagnose wird hierbei auch als Verdachts- oder Arbeitsdiagnose

bezeichnet und wird anhand neuer Informationen fortlaufend überprüft und gegebenenfalls aktualisiert. Die zentrale Frage während der zyklischen Überprüfung der Verdachtsdiagnose ist, ob genügend Informationen gesammelt wurden, um sich auf die Diagnose festzulegen und mit der Behandlung zu beginnen. Trotz modernster Technologie und langjähriger Erfahrung bringt jede Diagnose eine gewisse Unsicherheit mit sich. Diese Tatsache wird von Jerome P. Kassirer in seinem Artikel über das Streben nach diagnostischer Gewissheit beschrieben: „Absolute certainty in diagnosis is unattainable, no matter how much information we gather, how many observations we make, or how many tests we perform. [...] Our task is not to attain certainty, but rather to reduce the level of diagnostic uncertainty enough to make optimal therapeutic decisions“ [34, S. 1489]. Um die bestmögliche Versorgung des Patienten zu gewährleisten, muss die Ungewissheit der Verdachtsdiagnose also nicht eliminiert, sondern ausreichend minimiert werden.

Nachdem sich das medizinische Personal auf eine Diagnose festgelegt hat, wird diese mit dem Patienten kommuniziert und mit der Behandlung begonnen. Auch während des Behandlungsprozesses ist die Sammlung und Interpretation von Informationen nicht abgeschlossen. Wie in Abbildung 3.1 am roten Pfeil zu erkennen ist, werden neue Erkenntnisse kontinuierlich ausgewertet, um die die Effektivität der Behandlung zu überprüfen. Bei einem Patienten mit einer Pneumonie könnten beispielsweise regelmäßig weitere Röntgenbilder aufgenommen werden. Während dieser Phase könnte das entwickelte Assistenzsystem für eine Entlastung der Fachärzte sorgen, indem es die Röntgenbilder fortlaufend analysiert und so das Abheilen der Entzündung sicherstellt. Nachdem sich der gesundheitliche Zustand des Patienten ausreichend verbessert hat, wird die Behandlung abgeschlossen und der medizinische Diagnoseprozess beendet.

3.2 Assistenzsysteme

Im Gesundheitssektor dienen Assistenzsysteme der Unterstützung des Fachpersonals in bestimmten Situationen und Abläufen. Diese interdisziplinären Systeme beinhalten unter Anderem Elemente aus den Bereichen Künstliche Intelligenz (KI), maschinelles Sehen, Robotik und Automatisierungstechnik. Die Anwendungsbereiche sind sehr vielfältig und reichen von der automatisierten Überwachung von Vital-Parametern, über Roboter-gestützte Rehabilitation nach einem Schlaganfall, bis hin zu CAD-Systemen zur Diagnose von Krankheiten wie in dieser Arbeit. CAD-Systeme erfüllen dabei stets eine unterstützende Rolle und sind nicht dafür gedacht, medizinisches Fachpersonal zu ersetzen. Auch wenn Diagnosen und Handlungsempfehlungen durch das System vorgeschlagen werden, liegt die endgültige Entscheidung für eine Diagnose oder Behandlung stets beim Arzt, da dieser auch die Verantwortung trägt.

Die computergestützte Diagnose von Krankheiten in Röntgenbildern ist im Vergleich zu anderen Assistenzsystemen eine noch sehr junge Technologie. Im Jahr 2018 wurde das Produkt qXR der Firma Qure.ai [63] weltweit zum ersten KI-basierten Produkt zur Erkennung von Krankheiten in Thoraxröntgenbildern, das eine CE-Zertifizierung erhalten hat. Da die Software und die darunterliegenden Trainingsdaten kontinuierlich

erweitert werden, ist dieses CAD-System auch heute noch führend in diesem Bereich. Bei der ersten Veröffentlichung basierte das System auf 1,2 Millionen Röntgenbildern aus 27 Krankenhäusern, ehe die proprietäre Datenbank im Jahr 2019 auf 2,3 Millionen Bilder aus 45 Krankenhäusern erweitert wurde [62]. Der wissenschaftliche Artikel von Putha u. a. wurde seitdem nicht mehr aktualisiert, jedoch wird im Internetauftritt von Qure.ai zum Zeitpunkt dieser Arbeit angegeben, dass die Datenbank aktuell 3,7 Millionen Thoraxröntgenbilder enthält [63]. Ebenso wie diese Arbeit verwendet qXR ausschließlich frontale Röntgenbilder (siehe Abschnitt 3.4.2).

Ein besonders nennenswertes nicht-kommerzielles Projekt dieses Bereichs ist Chester von Cohen u. a. [13]. Chester verwendet größtenteils die gleichen Datensätze wie diese Arbeit und ist frei im Internet abrufbar. Die Intention des Systems ist es, das Potential Deep Learning-basierter Diagnosen für die Allgemeinheit verfügbar zu machen. Das Modell inklusive seiner Gewichte wird dafür im Browser geladen, um die Vorhersagen direkt auf der Hardware des Nutzers durchzuführen (siehe Abschnitt 6.4.2).

3.3 Diagnose von Pneumonien

Da der allgemeine Diagnoseprozess nun bekannt ist, soll im folgenden Abschnitt gezielt auf die Diagnose von Pneumonien anhand von Röntgenbildern eingegangen werden. Ziel ist es, ein Grundverständnis für die Aufnahme der Bilder, ihrer Beschreibung und den charakteristischen Anzeichen der Krankheit aufzubauen. Das Erkennen und Beurteilen dieser Merkmale in Röntgenbildern ist die Hauptaufgabe des in Kapitel 5 beschriebenen Vorhersagemodells.

3.3.1 Die Thoraxaufnahme

Im medizinischen Kontext wird der menschliche Brustkorb auch als Thorax bezeichnet. Die Röntgenaufnahme des Thorax ist eine Routine-Untersuchung, die verwendet wird, um den Zustand des knöchernen Brustkorbs und der inneren Organe zu beurteilen. Sowohl für akute als auch für chronische Erkrankungen liefern die Bilder entscheidende diagnostische Informationen und werden in fast allen medizinischen Disziplinen eingesetzt. Die vielfältige Einsetzbarkeit macht die Thoraxaufnahme im Vergleich zu anderen Röntgenpraktiken besonders. So wird die Befundung von Thoraxaufnahmen abhängig von der diagnostischen Fragestellung von Ärzten unterschiedlicher Fachrichtungen durchgeführt. Im Kontrast dazu werden beispielsweise Röntgenaufnahmen der Füße, Knie und Beine hauptsächlich von Orthopäden angeordnet und befundet.

Die Heterogenität der untersuchenden Ärzte ist einer der Gründe weshalb keine standardisierte, sondern eine fachspezifische Befundung stattfindet. Grundsätzlich wird bei der Analyse der Bilder jedoch immer mit sogenannten Röntgenzeichen gearbeitet. Unter Röntgenzeichen versteht man charakteristische Merkmale in Röntgenbildern, die auf bestimmte Erkrankungen hinweisen können. Diese Merkmale können in Thoraxaufnahmen beispielsweise in Form von Aufhellungen, Schatten, Struktur-, Form-

oder Lageveränderungen auftreten. Die Zeichen sind dabei oftmals so typisch für eine bestimmte Erkrankung, dass sie gemeinsam mit den ausformulierten Befunden verwendet werden, um eine Diagnose zu beweisen. Auch bei der Diagnose von Pneumonien werden diese Veränderungen auf Röntgenbildern zurate gezogen, was nachfolgend erläutert wird.

3.3.2 Röntgenzeichen einer Pneumonie

In diesem Abschnitt soll erklärt werden, wie sich eine Pneumonie auf Röntgenbildern äußert. Das wohl deutlichste Anzeichen für eine Pneumonie ist das Auftreten von Schatten in der Lungenregion. Diese können sich in unterschiedlichen Bereichen der Lunge befinden, einseitig oder symmetrisch auftreten und in ihren Dimensionen variieren.

Entgegengesetzt zum allgemeinen Sprachgebrauch bezeichnen Verschattungen in der medizinischen Bildgebung helle bis weiße Bereiche. Ebenso wird auch der Begriff Aufhellung konträr zum natürlichen Sprachgebrauch verwendet und bezeichnen dunkle bis schwarze Auffälligkeiten in Röntgenbildern. Diese Terminologie lässt sich auf die Frühzeit der Röntgentechnik zurückführen. Wie bereits in Abschnitt 2.2.3 beschrieben wurde, erschienen auf den ersten Röntgenaufnahmen strahlenundurchlässige Gewebe und Objekte auf den Bildern als dunkle Bereiche. Daher wurden Regionen, die dunkler waren als erwartet als Verschattung bezeichnet. Analog dazu wurden Regionen, die heller waren als erwartet als Aufhellung bezeichnet. Seit der Erfindung des Röntgenfilms sind diese Darstellungen jedoch invertiert. Da die Beschreibung der Strahlendurchlässigkeit durch die erwähnten Begriffe zu diesem Zeitpunkt bereits weit verbreitet war, wurden die Begriffe trotz ihrer nun visuell entgegengesetzten Darstellung beibehalten.

Die Abbildungen 3.2, 3.3 und 3.4 zeigen Beispiele des Erscheinungsbilds von Pneumonien in Röntgenbildern. Alle Bilder und die dazugehörige Deutung von Röntgenzeichen stammen aus dem Buch „Röntgendiagnostik von Thoraxerkrankungen“ von Prof. Dr. med. H. M. Kulke [38]. Abbildung 3.2 zeigt eine junge Patientin, die seit zwei Tagen über geringen Husten klagt. Das Röntgenbild zeigt auf der rechten Seite einen mitteldichten, überwiegend homogenen Schatten mit scharfen Rand nach oben und einem unscharfen Rand nach unten. Aufgrund der im vorherigen Abschnitt beschriebenen Konvention zur Beschreibung der Seite, befindet sich die als rechts bezeichnete Seite in der linken Bildhälfte. Diese Festlegung sollte ebenso wie die zuvor beschriebene Definition von Verschattung und Aufhellung beim Lesen dieser Arbeit berücksichtigt werden. Der Arzt vermutet eine sogenannte Mittellappenpneumonie, welche er als aktuelle Verdachtsdiagnose festlegt (siehe Abschnitt 3.1). Da er anhand des Röntgenbilds nicht ausschließen kann, dass es sich bei der Verschattung um einen Tumor handelt, ordnet er umgehend eine Bronchoskopie und eine CT-Untersuchung an. Nachdem der CT-Befund vorliegt, konnte die diagnostische Unsicherheit (siehe Abschnitt 3.1) ausreichend minimiert und sich auf die Diagnose der Mittellappenpneumonie festgelegt werden.

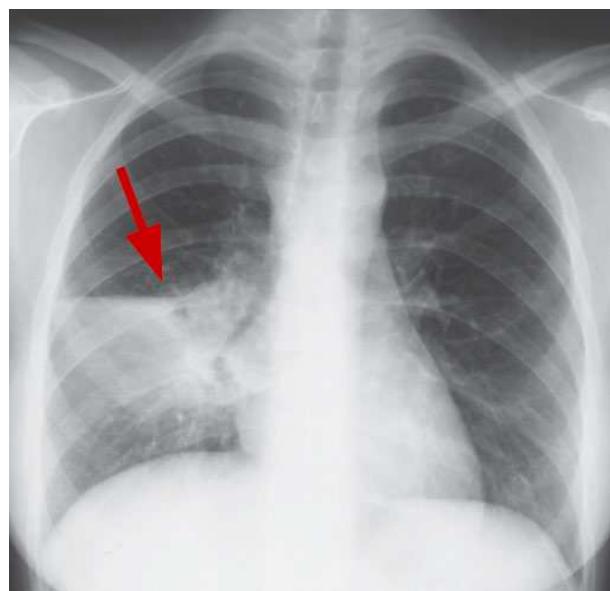


Abbildung 3.2: Pneumonie im mittleren Bereich der rechten Seite des Thorax, erkennbar am homogen auftretenden Schatten. (Quelle: [38, S. 63])

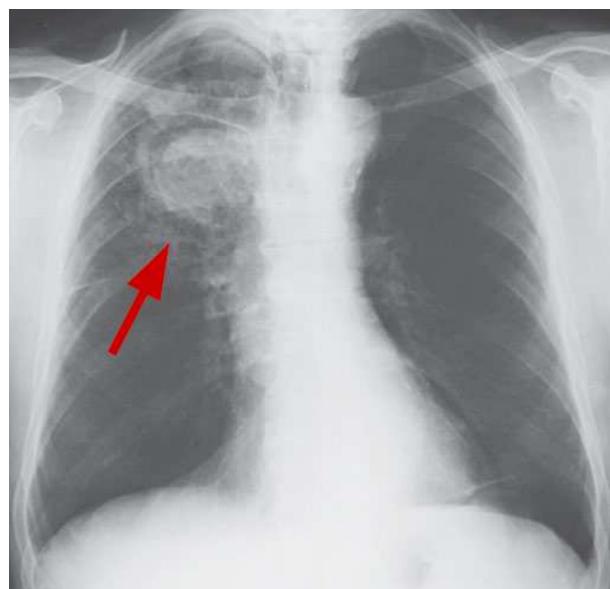
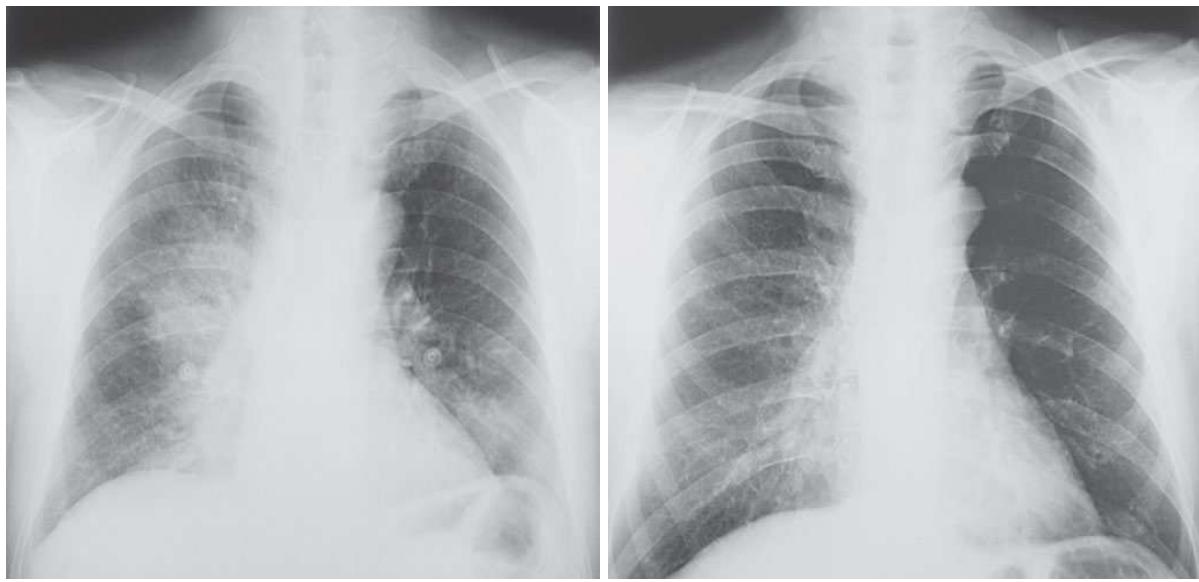


Abbildung 3.3: Pneumonie im rechten oberen Bereich, erkennbar am runden Schatten. Die längliche Aufhellung innerhalb des Schattens deutet auf eine weitere Infektion hin, die unabhängig von der Pneumonie ist. (Quelle: [38, S. 72])



(a) Akute zentrale Pneumonie

(b) Abgeheilte zentrale Pneumonie

Abbildung 3.4: Vergleich der Röntgenbilder eines Patienten mit zentraler Pneumonie vor (3.4a) und 14 Tage nach (3.4b) der medikamentösen Behandlung. (Quelle: [38, S. 65f])

Abbildung 3.3 zeigt einen Patienten, bei dem eine bakterielle Pneumonie bereits vor 10 Tagen diagnostiziert wurde. Die Röntgenaufnahme zeigt, dass die bereits diagnostizierte Pneumonie im rechten oberen Bereich sich durch die medikamentöse Therapie etwas zurückgebildet hat. Die Verschattung wird vom Arzt als mitteldichter inhomogener Schatten mit unscharfem Rand beschrieben, in dessen Mitte sich ein bogenförmig verlaufender Aufhellungsstreifen mit scharfem innerem Rand befindet. Aufgrund der länglichen Aufhellung innerhalb des Schattens äußert der Arzt den Verdacht auf eine weitere, von der Pneumonie unabhängige, Infektion. Kommt wie in dieser Situation ein weiterer Erreger zu einer bestehenden Entzündung hinzu, spricht man auch von einer Superinfektion [vgl. 38, S. 72].

Um den direkten Vergleich einer infizierten Lunge mit einer gesunden Lunge zu ermöglichen, werden in Abbildung 3.4 zwei Röntgenbilder des gleichen Patienten gegenübergestellt. Die Aufnahmen zeigen einen 56-jährigen Patienten, der seit einigen Wochen über Husten, Fieber und weitere Symptome klagt. Im Gegensatz zu den vorherigen Beispielen hat die Entzündung größere Teile der Lunge befallen und befindet sich in beiden Lungenflügeln. Der Arzt stellt in Abbildung 3.4a einen homogenen Schatten mit scharfem Rand im zentralen Bereich, sowie inhomogene Schatten mit vereinzelten streifigen Ausziehungen und Fleckschatten im rechten mittleren und linken unteren Bereich fest. Das Erscheinungsbild deutet auf eine sogenannte zentrale Pneumonie mit rechtsseitiger Ausprägung hin, welche die vorläufige Verdachtsdiagnose darstellt. Ähnlich wie in Abbildung 3.2 kann auch hier Lungenkrebs als Krankheitsursache durch das Röntgenbild nicht ausgeschlossen werden. Folglich wird auch hier umgehend eine Bronchoskopie und CT-Aufnahme angeordnet. Nachdem der CT-Befund die Verdachtsdiagnose bestärkt, wird mit der medikamentösen Behandlung begonnen. In Abbildung 3.4b ist die Thoraxaufnahme des Patienten nach 14-tägiger Behandlung der Lungenentzündung zu sehen. Der gesundheitliche Zustand des Patienten hat sich

seitdem deutlich verbessert. Es ist eindeutig zu erkennen, dass sich die Verschattungen in allen Bereichen zurückgebildet haben. Die Diagnose und Therapie der Pneumonie ist somit allumfänglich gelungen.

3.3.3 Computertomographie zur Pneumoniediagnose

Die Röntgenaufnahme des Brustkorbs reicht in der Regel aus, um die Diagnose einer Lungenentzündung zu bestätigen und die Wirksamkeit der Behandlung zu beurteilen. In manchen Fällen kann es jedoch von Vorteil sein zusätzlich eine CT-Aufnahme einzusetzen. Dies ist vor allem sinnvoll, wenn weiterführende Analysen notwendig sind, um die Pneumonie zu diagnostizieren, oder der Verdacht auf eine andere, im Röntgenbild nicht klar erkennbare, Krankheit im Raum steht.

So ist es beispielsweise möglich, CT-Aufnahmen bei immungeschwächten Patienten einzusetzen, um eine Pneumonie diagnostizieren zu können [vgl. 50, S. 780]. Bei besonders komplexen Pneumonien werden häufig laterale Röntgenaufnahmen (siehe Abschnitt 3.4.2) zusätzlich zur frontalen Röntgenaufnahme eingesetzt. Ist auch die Kombination der beiden Röntgenbilder nicht aussagekräftig genug, kann eine CT-Aufnahme Klarheit über die Lokalisation und Beschaffenheit der potenziellen Pneumonie verschaffen.

Auch wenn der klinische Verdacht auf eine Pneumonie hindeutet, kann das Röntgenbild im Kontrast dazu stehen und normal ausfallen. Unter klinischem Verdacht oder klinischem Befund versteht man die Feststellung einer Krankheit aufgrund der Symptome, Krankenhistorie und körperlichen Untersuchung des Patienten. In dieser Situation würden aufgrund der widersprüchlichen Befunde Zweifel an der Unauffälligkeit des Röntgenbefunds auftreten. Hier kann eine CT-Aufnahme den Ärzten als zusätzliche Informationsquelle dienen und gegebenenfalls den Röntgenbefund bestärken oder entkräften.

Im vorherigen Abschnitt wurden bereits Beispiele vorgestellt, in denen der Röntgenbefund nicht eindeutig war (siehe Abbildung 3.2 und 3.4a). In beiden Fällen gab es durch charakteristische Merkmale im Röntgenbild Indizien für einen möglichen Tumor. Diese sind in Röntgenbildern zu sehen, jedoch nicht zweifelsfrei identifizierbar. Durch eine zusätzliche CT-Untersuchung kann in diesen Situationen Klarheit über die Verdachtsdiagnose gewonnen werden.

3.4 Perspektivische Eigenschaften von Röntgenbildern

Kapitel 4 diskutiert die Eigenschaften mehrerer öffentlich verfügbarer Datensätze für Thoraxröntgenbilder. Unter anderem werden dabei die Richtungsbezeichnungen und Perspektiven der Bilder mit medizinischer Fachsprache beschrieben. Daher erläutert dieser Abschnitt die zur Analyse der Datensätze notwendigen Vorkenntnisse über perspektivische Eigenschaften von Röntgenbildern.

3.4.1 Lage- und Richtungsbezeichnungen

Im Bereich der Anatomie wird eine Nomenklatur verwendet, die es ermöglicht Körperteile, Regionen und Lagebeziehungen unmissverständlich lokalisieren und kommunizieren zu können. Diese ist für Röntgenaufnahmen besonders wichtig, damit der Röntgentechniker oder Arzt mitteilen kann, in welcher Perspektive ein Bild aufgezeichnet wurde (siehe Abschnitt 3.4.2). Abbildung 3.5 zeigt eine Übersicht dieser Richtungsbezeichnungen des menschlichen Körpers. Speziell die Begriffe Anterior, Posterior und Lateral finden häufige Verwendung in der Radiologie und sind für diese Arbeit relevant. Anterior bezeichnet die Brust- und Bauchseite des Körpers, welche umgangssprachlich vorne genannt wird. Die gegenüberliegende Seite, auf der sich der Rücken befindet, wird als Posterior oder umgangssprachlich hinten bezeichnet. Unter Lateral versteht man die Seitenansicht auf den menschlichen Körper.

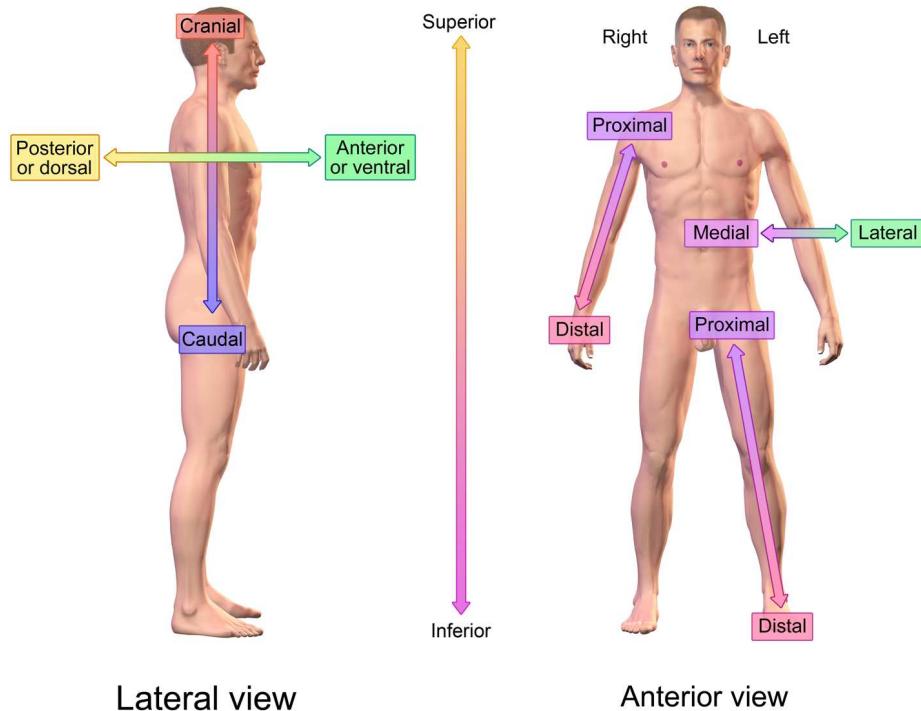


Abbildung 3.5: Anatomische Richtungen des Menschen in medizinischer Fachsprache (Quelle: [8])

Da Röntgenaufnahmen von beiden Seiten betrachtet werden können und der menschliche Körper überwiegend symmetrisch aufgebaut ist, wird im Röntgenbild nicht immer klar ersichtlich, welche Seite links oder rechts ist. Sollte eine Aufnahme spiegelverkehrt interpretiert werden, kann dies im schlimmsten Fall zu Fehldiagnosen und der falschen Behandlung einer Krankheit führen. Um Klarheit über die sogenannte Lateralität zu erhalten, werden daher häufig Markierungen wie beispielsweise „R“, „L“ oder „exposure side“ auf den Röntgenbildern platziert. Die Konvention zur Seitenbezeichnung ist nicht nur für die Bilder selbst, sondern auch für die Beschreibung der Röntgenaufnahmen während der Befundung besonders wichtig. Daher wird konsequent die Seite des

Körpers referenziert, wenn von links beziehungsweise rechts gesprochen wird (siehe Abbildung 3.5). Bei der Befundung einer Thoraxaufnahme würde der Arzt somit beispielsweise vermerken, dass sich eine Verletzung auf der rechten Seite befindet, obwohl die Verletzung sich in der linken Bildhälfte befindet. Da sich das Herz asymmetrisch auf der linken Seite des Oberkörpers befindet und gut auf Röntgenbildern sichtbar ist, kommt es bei Thoraxaufnahmen nur äußerst selten zur Verwechslung der Seite [vgl. 19].

3.4.2 Bildperspektiven

Abhängig von der diagnostischen Fragestellung können unterschiedliche Bildperspektiven von Vorteil sein. Die Perspektive eines Röntgenbilds, auch Projektion genannt, setzt sich aus mehreren Eigenschaften und Gegebenheiten zum Zeitpunkt der Bildaufnahme zusammen. In Tabelle 3.1 ist eine Übersicht der wichtigsten Eigenschaften von Thoraxaufnahmen zu sehen. Unter anderem beinhalten perspektivische Eigenschaften Informationen darüber, wie das Röntgengerät und der Patient während der Aufnahme positioniert wurden. Bei der Befundung der Bilder sind diese Informationen für den Arzt von zentraler Bedeutung, da sie die Bilder in einen anderen Kontext rücken. Für eine vollständige Beschreibung der Projektion müssen die Ansicht, Strahlenrichtung und Lage des Patienten angegeben werden (siehe Tabelle 3.1).

Für die meisten diagnostischen Fragestellungen im Brustbereich liefert eine Aufnahme in der Perspektive Frontal PA Erect die besten Ergebnisse. Der Patient steht dabei aufrecht mit der Brust zum Detektor gerichtet (siehe Abbildung 3.6a). Durch die frontale Ansicht kann die größtmögliche Fläche der Lunge visualisiert werden und eine Inspektion beider Lungenflügel stattfinden. Die Ausrichtung PA sorgt dafür, dass sich das Herz möglichst nah am Detektor befindet, damit die Proportionen des Organs nicht verzerrt werden (siehe Abbildung 3.7). Aufgrund der Schwerkraft ist es vorteilhaft, wenn sich der Patient in aufrechter Lage (Erect) befindet, da dies die inneren Organe in ihrer normalen Form und Position auf dem Röntgenbild erscheinen lässt.

Die Ansicht Lateral (siehe Abbildung 3.6b) wird in der Regel ergänzend eingesetzt, um Bereiche besser sichtbar zu machen, die bei der zuvor beschriebenen Standardperspektive durch Organe oder Knochen verdeckt werden. Die seitliche Ansicht des Thorax kann zudem bei der Lokalisierung von Fremdkörpern und Entzündungen hilfreich sein, da durch die frontale Ansicht nicht ersichtlich wird, wie tief die Verschattungen beziehungsweise Aufhellungen im Körper positioniert sind.

In lordotischer Lage formt der Oberkörper des Patienten ein Hohlkreuz, statt flach auf dem Detektor aufzuliegen (siehe Abbildung 3.6c). Aufgrund des Winkels kann diese Aufnahmetechnik verwendet werden, um verdächtige Bereiche innerhalb der oberen Lungenspitzen zu beurteilen, die bei regulären Thoraxaufnahmen durch darüber liegendes Gewebe und Knochen verdeckt werden [vgl. 26, S. 360]. Da die räumliche

Eigenschaft	Bedeutung	Erklärung
Ansicht		
Frontal	Vorderseite	Der Oberkörper des Patienten ist entweder zur Röntgenröhre oder zum Detektor gerichtet.
Lateral	Seitlich	Der Patienten steht seitlich, sodass der Oberkörper senkrecht zur Röhre und zum Detektor steht.
Strahlenrichtung		
Postero-Anterior (PA)	Von hinten nach vorne	Die Röntgenstrahlen treffen zuerst auf die Seite des Patienten, die weiter vom Herz entfernt ist.
Antero-Posterior (AP)	Von vorne nach hinten	Die Röntgenstrahlen treffen zuerst auf die Seite des Patienten, die näher am Herz ist.
Lage des Patienten		
Erect / Vertikal	Aufrecht	Der Patient befindet sich entweder stehend oder sitzend in aufrechter Position.
Supine / Horizontal	Liegend	Der Patient befindet sich in liegender Position während der Aufnahme.
Lordotisch	Geneigt	Der Patient neigt den Oberkörper während der Aufnahme nach hinten, sodass die Röntgenstrahlen leicht angewinkelt von unten auf den Oberkörper treffen.

Tabelle 3.1: Perspektivische Eigenschaften von Thoraxaufnahmen

Darstellung von CT-Aufnahmen der lordotischen Perspektive überlegen ist und die Verbreitung von CT-Geräten fortlaufend zunimmt, geht die Nutzung dieser Projektionsart zunehmend zurück.

Sollte es der gesundheitliche Zustand des Patienten nicht ermöglichen, eine sitzende oder stehende PA-Aufnahme durchzuführen, kann eine Röntgenaufnahme im Liegen angefertigt werden. In der Regel befindet sich der Patient dabei auf dem Rücken liegend (siehe Abbildung 3.6e). Eine Aufnahme in Bauchlage ist aufgrund des kritischen Zustands bei bettlägerigen Patienten nicht möglich [vgl. 24, S. 24]. Man spricht dann auch von einer Frontal AP Supine Projektion. Abgesehen von der einfachen Handhabung der Lagerung des Patienten bringt diese Vorgehensweise jedoch eine Reihe an Nachteilen mit sich. Da sich die Röntgenstrahlung kegelförmig ausbreitet und das Herz durch die AP-Ausrichtung nun weiter vom Detektor entfernt ist, erscheint der Schatten des Herzens größer als auf einer PA-Aufnahme (siehe Abbildung 3.7). Das größer erscheinende Herz sorgt dafür, dass der dazugehörige Schatten größere Teile des Röntgenbilds bedeckt und in diesen Bereichen dadurch die Bildqualität mindert.

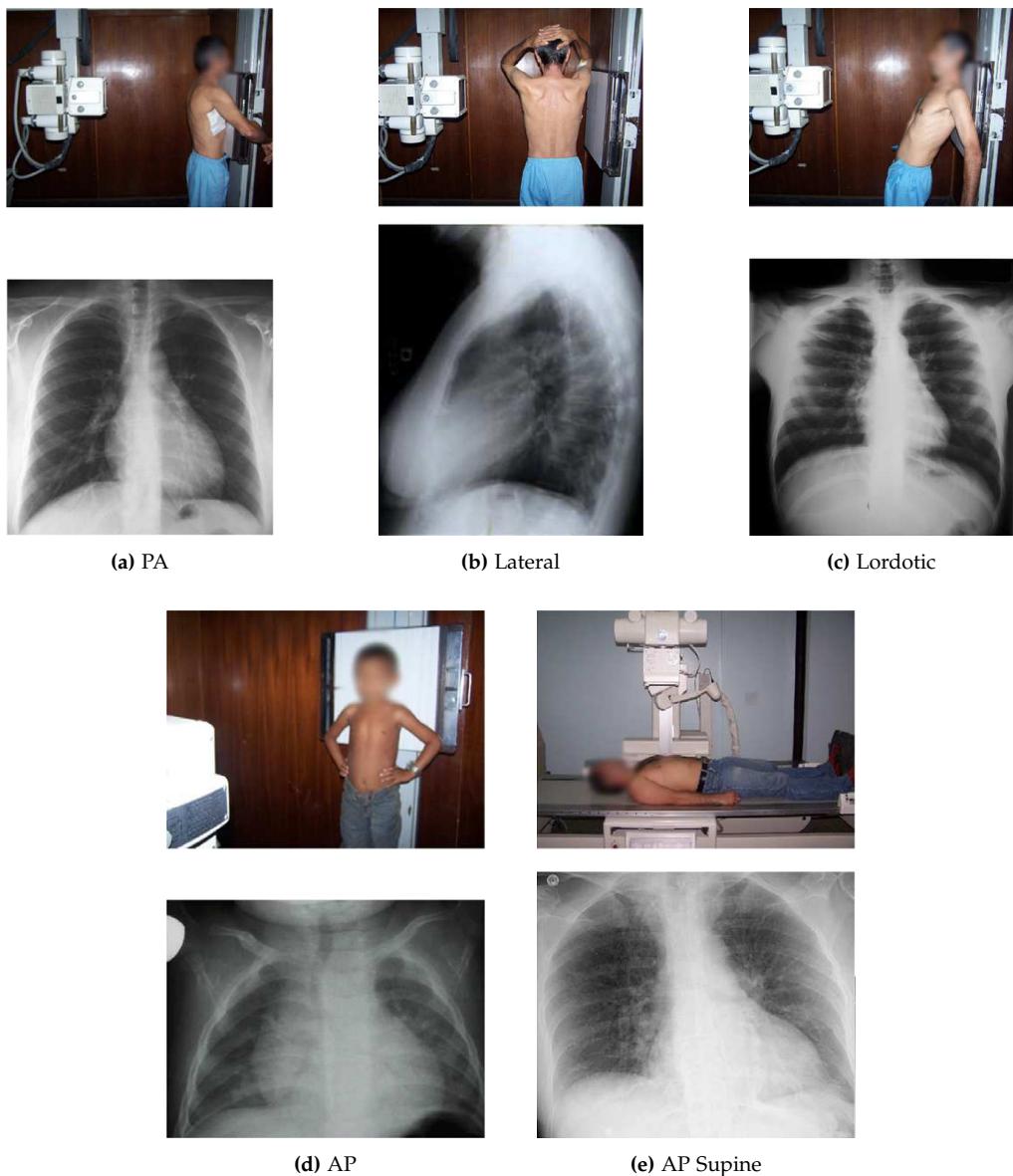


Abbildung 3.6: Übersicht verbreiteter Perspektiven von Thorax Röntgenaufnahmen (Quelle: [10, S. 5])

Dieses Verhalten wird in Abbildung 3.6d durch die besonders große Silhouette des Herzens deutlich. Des Weiteren sorgt die Gravitation in liegender beziehungsweise jeder nicht-aufrechten Position dafür, dass Gefäße und Organe verzerrt dargestellt werden. Die horizontale AP-Aufnahme nimmt somit eine schlechtere Bildqualität in Kauf, um Patienten in meist kritischem gesundheitlichem Zustand zu entlasten.

In Abschnitt 3.4.1 wurde bereits erwähnt, dass sich das Herz in der linken Körperhälfte des Menschen befindet. Bei einer PA-Aufnahme ist das Herz in der rechten Bildhälfte zu sehen, so als würde man dem Patienten gegenüberstehen und durch ihn blicken. Die Perspektive einer AP-Aufnahme ist zum Zeitpunkt der Aufnahme umgekehrt zur PA-Aufnahme, als würde man hinter dem Patienten stehen. Per Konvention werden frontale Thoraxaufnahmen jedoch so dargestellt, als wären sie in PA-Ansicht aufgenommen worden, da dies die Standardperspektive ist. Das bedeutet, dass Rönt-

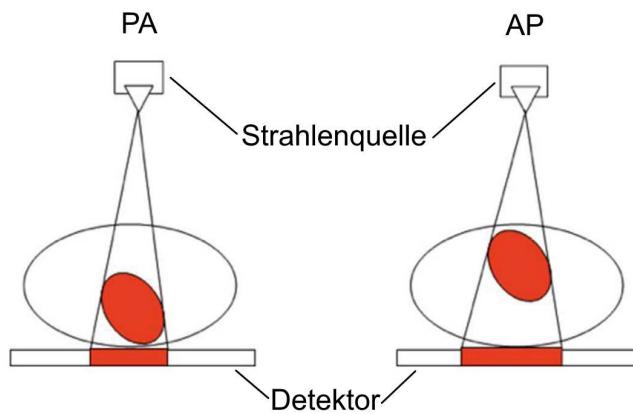


Abbildung 3.7: Schematischer Vergleich der Größe des Herzens bei PA- und AP-Aufnahmen. Die Größe auf dem Röntgenbild (rotes Rechteck) hängt von der Distanz des Herzens (rote Ellipse) zur Strahlenquelle und zum Detektor ab. (Quelle: [10, S. 6])

genbilder in AP-Perspektive gespiegelt werden, bevor sie betrachtet werden. Dadurch sind die Bilder trotz ihrer unterschiedlichen Aufnahmetechniken vergleichbarer und das Herz befindet sich immer in der rechten Bildhälfte.

Kapitel 4

Auswahl und Modellierung der Daten

Das maschinelle Sehen hat durch die stetig steigende Leistungsfähigkeit von Computerhardware in Kombination mit neuen CNN-Architekturen rasante Fortschritte erreicht. Bis vor einigen Jahren konnten diese Fortschritte aufgrund des Mangels an Trainingsdaten jedoch nicht auf medizinische Problemstellungen angewendet werden, was auch die Autoren des in Abschnitt 4.2.2 beschriebenen Datensatzes erkannten:

The main limitation is that all proposed methods are evaluated on some small-to-middle scale problems of (at most) several hundred patients. It remains unclear how well the current deep learning techniques will scale up to tens of thousands of patient studies. [84, S. 2]

Seit dieser Aussage aus dem Jahr 2017 wurde eine Vielzahl medizinischer Datensätze veröffentlicht, weshalb sich Technologien des maschinellen Sehens auch zunehmend im Gesundheitssektor etablieren konnten. Dieses Kapitel analysiert mehrere Datensätze für Thoraxröntgenbilder, welche die Basis für das entwickelte Vorhersagemodell bilden und ist daher ein sehr bedeutender Teil dieser Arbeit. Der erste Abschnitt des Kapitels beschreibt die zugrundeliegende Problematik unausgeglichenener medizinischer Datensätze beim Einsatz von neuronalen Netzen. Motiviert durch die vorangegangene Problembeschreibung führt Abschnitt 4.2 eine sehr ausführliche Analyse und Bereinigung der Datensätze durch. Im letzten Abschnitt werden Konzepte zur Verbesserung der Daten entwickelt und abschließend der erzeugte kombinierte Datensatz vorgestellt.

4.1 Einfluss der Eigenschaften

Zusätzlich zu den Diagnosen enthalten medizinische Datensätze häufig auch Metainformationen, wie beispielsweise das Alter oder Geschlecht der Patienten. Die Informationen variieren hierbei sehr stark, was unter anderem auf die verschiedenen Datenschutzrichtlinien der Krankenhäuser zurückzuführen ist, in denen die Bilder aufgenommen wurden. Die Verteilung dieser Eigenschaften ist nicht immer gleichmäßig, was negative Folgen auf das Training neuronaler Netze haben kann. Im Bereich des maschinellen Lernens wird diese Voreingenommenheit auch als Bias bezeichnet.

Die Auswirkungen ungleich verteilter Eigenschaften auf das Modelltraining wurden von Seyyed-Kalantari u. a. speziell für Thoraxröntgenbilder näher untersucht [75]. In ihrem Artikel kritisieren sie die Verteilung der Eigenschaften in den Datensätzen ChestX-ray14 (CXR14), CheXpert und MIMIC-CXR, welche in dieser Arbeit ebenfalls verwendet werden. Dabei konnten sie nachweisen, dass sich der Bias der Daten auf die Vorhersagen des Modells überträgt und Patienten beispielsweise aufgrund ihres Geschlechts benachteiligt behandelt wurden. Das neuronale Netz erlernt und reproduziert somit systematische Diskriminierungen, welche sich in den Daten befinden. Die Ergebnisse zeigen, wie dringlich die Berücksichtigung der Metainformationen bei der Entwicklung von CAD-Systemen ist. Die Eigenschaften der Patienten werden bei der nachfolgenden Analyse der Datensätze daher besonders fokussiert. Weiterhin werden die Unterschiede der Eigenschaften von Patienten mit und ohne Pneumonie untersucht, um den Einfluss der Eigenschaften auf die Vorhersage der Krankheit zu minimieren.

4.2 Analyse und Bereinigung der Datensätze

Insgesamt verwendet diese Arbeit sechs öffentlich verfügbare Datensätze verschiedener Institutionen. Dieser Abschnitt analysiert die Eigenschaften und Qualität der Daten, um ein Verständnis für die Datenlage zu entwickeln und anschließend ungeeignete Daten zu entfernen. Das Wissen über die Datensätze bietet die Grundlage für die in Abschnitt 4.3 angewendete Stratifikation und Balancierung der Klassen. Das CAD-System soll Pneumonien von gesunden Lungen unterscheiden. Daher sind nur Patientenstudien relevant, die diese Fälle abdecken und entweder keine Krankheiten, oder eine Pneumonie vorweisen. Die Präsenz anderer Krankheiten zusätzlich zur Pneumonie ist dabei zulässig, da sonst die Anzahl geeigneter Daten zu gering wäre. Manche der vorliegenden Datensätze beinhalten ergänzend zu den regulären frontalen Röntgenbildern auch laterale Bilder, um zusätzliche Informationen bereitzustellen. Im Rahmen dieser Arbeit ist die Entwicklung mehrerer Vorhersagemodelle für verschiedene Ansichten nicht vorgesehen. Die nachfolgende Bereinigung der Datensätze entfernt daher laterale Bilder.

Für jeden Datensatz wurde ein Jupyter Notebook entwickelt, welches die Metainformationen der Datensätze einliest, analysiert und filtert. Abbildung 4.1 visualisiert den Ablauf der Analyse. Eine Einheitliche Untersuchung der unterschiedlichen Datensätze wird durch das zuvor implementierte Python-Modul `dataset_analyzer` ermöglicht. Das Modul kapselt diverse wiederkehrende Aufgaben und wird in der Präambel der Jupyter Notebooks importiert. Die Ergebnisse der Datenanalyse und -auswahl befinden sich in einem pandas Dataframe, welcher am Ende der Analyse im Dateiformat Comma-Separated Values (CSV) exportiert und abgespeichert wird.

Die Datensätze sind nachfolgend chronologisch anhand des Jahres der Veröffentlichung sortiert. Auf diese Weise kann ein Einblick in die fortschreitende Entwicklung und zunehmende Komplexität der Datensätze gewonnen werden. Die nachfolgend

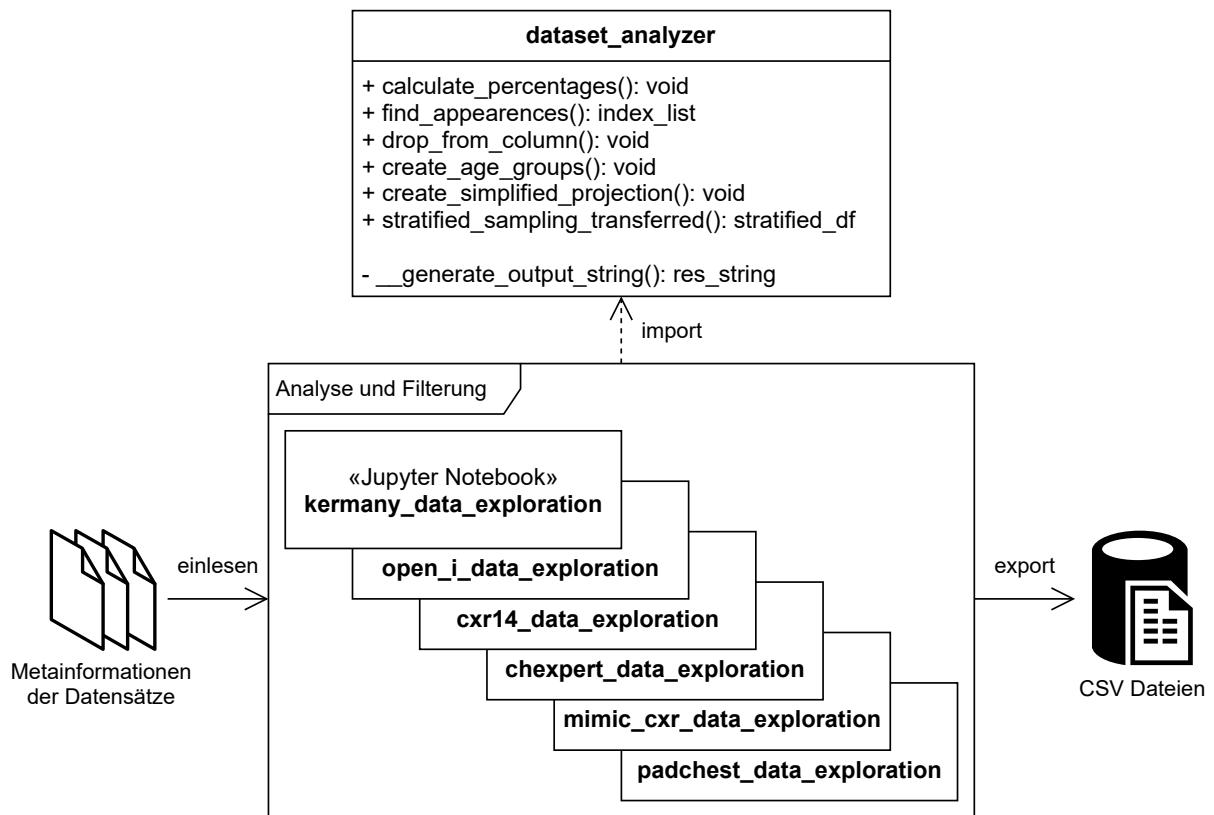


Abbildung 4.1: Analyse und Filterung der Datensätze durch Jupyter Notebooks

aufgelisteten Kennzahlen sind in jedem Abschnitt in tabellarischer Form zu finden, um eine schnelle und einfache Übersicht zu ermöglichen.

- Herausgeber
- Krankenhaus
- Anzahl Patienten
- Anzahl Röntgenbilder
- Labeling-Methode
- Verfügbarkeit

Alle verwendeten Datensätze enthalten ausschließlich anonymisierte Informationen, um die Privatsphäre der Patienten zu gewährleisten. Des Weiteren verpflichtet die Nutzung der Daten dazu, sie ausschließlich für nicht-kommerzielle Forschungs- und Bildungszwecke einzusetzen. Da im Rahmen dieser Arbeit keine kommerziellen Interessen vom Autor oder dem kooperierenden Unternehmen verfolgt werden, ist die rechtmäßige Nutzung der Daten gewährleistet. Die detaillierten Nutzungsrichtlinien des jeweiligen Datensatzes können aus den zitierten Veröffentlichungen entnommen werden.

4.2.1 Open-i

Open-i ist eine frei zugängliche Suchmaschine für biomedizinische Bilder der NLM [15]. Aktuell bietet das System Zugriff auf über 3,7 Millionen Bilder aus etwa 1,2 Millionen medizinischen Artikeln. Zusätzlich zu Bildern aus wissenschaftlichen Artikeln beinhaltet Open-i medizinische Bildersammlungen und Datensätze. Darunter befindet sich ein Datensatz für Thoraxröntgenaufnahmen und dazugehörigen Befunden, welcher im Jahr 2015 von der gleichen Autorin wie die Open-i Suchmaschine veröffentlicht wurde [16]. Ziel der Kollektion war es, eine öffentlich verfügbare und durchsuchbare Datenbank mit Röntgenbildern und dazugehörigen Befunden zu erstellen, um die Forschung, Entwicklung und Schulung in der Radiologie zu erleichtern. Zu diesem Zeitpunkt gab es kaum öffentlich verfügbare Sammlungen von medizinischen Bildern und Texten. Open-i war somit der erste große Datensatz im Radiologiebereich, der unter Einhaltung der Richtlinien für jeden frei zur Verfügung stand. Durch den unkomplizierten Zugang zu den Daten eröffneten sich sowohl für die Medizin als auch für die Informatik neue Möglichkeiten.

Eigenschaft	Wert
Herausgeber	National Library of Medicine
Krankenhaus	Zwei große Krankenhausssysteme aus dem Indiana Network for Patient Care ¹ , USA
Anzahl Patienten	3 955
Anzahl Röntgenbilder	7 470
Labeling-Methode	Manuell
Verfügbarkeit	Öffentlich zugänglich [53]

Tabelle 4.1: Übersicht des Open-i Datensatz

Bei der Auswahl der Patientenstudien wurde vorausgesetzt, dass mindestens ein Röntgenbild in PA-Projektion und eines in Lateral (L)-Projektion enthalten ist. Aus welchen Krankenhäusern die Bilder genau stammen, ist aus Datenschutzgründen nicht hinterlegt. Um eine manuelle Überprüfung der Daten zu ermöglichen, wurden lediglich 4 000 Studien mit 8 121 dazugehörigen Röntgenbildern verwendet. Durch die Identifizierbarkeit von Personen anhand von Schmuck, Zähnen und anderen Merkmalen musste ein Teil der Daten vor der Veröffentlichung entfernt werden (siehe Tabelle 4.1). Die Beurteilung der Studien wurde anhand der Radiologieberichte durchgeführt. Zwei in medizinischer Informatik ausgebildete Personen beurteilten unabhängig voneinander die Befunde in jedem Bericht. Bei unterschiedlichen Ergebnissen wurde der Radiologiebericht gemeinsam mit einer dritten geschulten Person diskutiert und ein Urteil gefällt. Durch die menschliche Kontrolle sind die Bilder des gesamten Datensatzes von sehr hoher Qualität. Im ersten Schritt wurden die Berichte in die Klassen *normal* und *not normal* unterteilt. Im zweiten Schritt wurden die Berichte der Klasse *not normal*

¹Das Indiana Network for Patient Care ist ein Netzwerk zum medizinischen Informationsaustausch und beinhaltet 38 Krankenhausssysteme mit insgesamt 117 Krankenhäusern [27].

in medizinische Unterkategorien klassifiziert. Hierbei wurden die Standards Medical Subject Headings (MeSH) [69] und Radiology Lexicon (RadLex) [40] verwendet, die zur Kategorisierung und Kodierung medizinischer Untersuchungsergebnisse eingesetzt werden. Die Definition und Bedeutung der MeSH und RadLex Codes können online unter [52], beziehungsweise [65] nachgeschlagen werden. Bei der Analyse der MeSH und RadLex Codes in dieser Arbeit zeigte sich, dass der Datensatz weniger als 200 Bilder von Patienten mit Pneumonie beinhaltet. Etwa die Hälfte der Bilder besteht zudem aus lateralen Röntgenbildern, welche für das hier entwickelte Vorhersagemodell irrelevant sind. Daher wurden aus dem Datensatz nur Bilder extrahiert, welche von gesunden Patienten stammen und in frontaler Ansicht aufgenommen wurden. Bei der Filterung anhand dieser Eigenschaften erwiesen sich 1387 frontale Röntgenbilder von 1379 Patienten als geeignet. Durch die Überrepräsentation von Patienten mit Pneumonie des in Abschnitt 4.2.3 beschriebenen Kermany Datensatz werden sich diese Bilder zu einem späteren Zeitpunkt der Arbeit noch als besonders nützlich erweisen (siehe Abschnitt 4.3.3).

4.2.2 ChestX-ray14

Wie in der Einleitung des Kapitels bereits erwähnt wurde, ist die Verfügbarkeit medizinischer Daten für die Entwicklung Deep Learning-basierter Vorhersagemodelle ein einschränkender Faktor. Diese Limitation war eine der Motivationen für die Entwicklung und Veröffentlichung von CXR14. Insgesamt beinhaltet der Datensatz 108 948 Bilder von 32 717 Patienten aus dem National Institutes of Health (NIH) Clinical Center und wurde im Jahr 2017 veröffentlicht (siehe Tabelle 4.2). Durch den Umfang des Datensatzes konnte ein bedeutender Schritt in Richtung Praxisrelevanz gemacht werden. Vermöglich gab es zu diesem Zeitpunkt bereits vergleichbare proprietäre Datensätze. Viele medizinische Datensätze bleiben der Öffentlichkeit jedoch vorenthalten, da es im Interesse der Unternehmen liegt, ihren Forschungs- und Wettbewerbsvorteil aufrecht zu erhalten (siehe Abschnitt 8.2).

Eigenschaft	Wert
Herausgeber	NIH Clinical Center
Krankenhaus	NIH Clinical Center, USA
Anzahl Patienten	32 717
Anzahl Röntgenbilder	108 948
Labeling-Methode	Kombination aus ML-basiert und regelbasiert
Verfügbarkeit	Öffentlich zugänglich [51]

Tabelle 4.2: Übersicht des CXR14 Datensatz

Crowdsourcing ist bei der Erstellung von großen klassifizierten Datensätzen ein beliebter Ansatz, um die Kosten gering zu halten. Diese Vorgehensweise ist für medizinische Daten gänzlich ungeeignet, da nur geschultes Personal in der Lage ist, Krankheiten

und Auffälligkeiten in den Bildern und Texten zu erkennen. Bereits beim zuvor beschriebenen Open-i Datensatz sorgte der hohe Aufwand bei der manuellen Beurteilung der Befunde für die Beschränkung auf 4 000 Radiologieberichte. Um eine weit höhere Anzahl an Berichten mit vertretbarem Aufwand beurteilen zu können, wurden anstelle der manuellen Vorgehensweise daher diverse Natural Language Processing (NLP) Techniken eingesetzt. Für die Detektion von Krankheiten wurde die Machine Learning Methode DNorm [42] und die Ontologie-basierte Methode MetaMap [4] verwendet und deren Ergebnisse zusammengeführt. Der entwickelte Algorithmus ist in der Lage die Erwähnung einer Krankheit aufgrund von Wörtern und Phrasen zu erkennen. Er ist jedoch nicht in der Lage Negationen und Unsicherheiten in den Sätzen zu detektieren. Daher wurden von den Forschen syntaktische Regeln entwickelt, die auf die Dependenzgrammatik¹ der Radiologieberichte angewandt wurden. Die Regeln der Autoren sind darauf ausgelegt, Negationen und Unsicherheiten bei der Erwähnung von Krankheiten zu erkennen. Konstrukte wie „es könnte eine Pneumonie vorliegen“ (Unsicherheit) oder „es liegt keine Pneumonie vor“ (Negation) würden vom System erkannt und ignoriert werden. Die Pipeline zur Detektion von Negationen und Unsicherheiten wurde separat unter dem Namen „NegBio“ veröffentlicht [60]. Zur Qualitätskontrolle des Systems wurde der NLP-Algorithmus auf die Radiologieberichte des im vorherigen Abschnitt erläuterten Open-i Datensatz angewandt. Die Ergebnisse wurden mit den manuellen Annotationen von Open-i verglichen, welche als Goldstandard verwendet wurden [vgl. 84, S. 4]. Hierbei erreichte die Methode insgesamt eine Genauigkeit von 0,90 und eine Sensitivität von 0,91.

Es gab zwar viel positive Resonanz für die Veröffentlichung der Daten, jedoch wurde die Qualität der Label von Ärzten sehr kritisch beäugt. Der Mediziner und zertifizierte Radiologe Luke Oakden-Rayner setzt sich im Rahmen seiner Forschung intensiv mit großen öffentlichen Datensätzen und den neuartigen Diagnosemöglichkeiten durch Deep Learning auseinander. Da die Radiologieberichte meist nicht Teil der Datensätze sind, führte er eine visuelle Analyse von je 700 Röntgenbildern aus den Datensätzen MURA [67] und CXR14 durch und verglich seine Ergebnisse mit den automatisiert erzeugten Labels der NLP-Systeme. Somit unterscheidet sich die zugrundeliegende Information zur Vergabe der Label zwischen den Datensätzen und Oakden-Rayner. Trotz dieser Einschränkung ist das Ergebnis seiner Studie hochgradig relevant, da den neuronalen Netzen bei der Klassifizierung von Bilddaten die dazugehörigen Texte ebenfalls nicht zur Verfügung stehen. Zusätzlich wurden Bilder festgehalten, bei denen es nicht eindeutig möglich war, das vergebene Label zu bestätigen oder zu dementieren. Aus diesen unklaren Fällen konnte anschließend eine kritische und eine kulante Beurteilung der Originalergebnisse erfolgen. Die Studie kam zu dem Resultat, dass die tatsächlichen Genauigkeiten der Krankheits-Label des CXR14 Datensatz zwischen 10 % (kulant) und 30 % (kritisch) unter den Werten der Originaldokumentation liegen. Bei als normal klassifizierten Bildern lagen die Genauigkeiten bei 0,76 (kulant), 0,62 (kritisch) und 0,87 (Originaldokumentation). Bilder, die laut CXR14 eine Pneumonie darstellen, erreichten in der Analyse eine Genauigkeit von 0,60 (kulant) und 0,50 (kritisch). Auch hier konnte der von Wang u. a. angegebene Wert von 0,66 nicht erreicht

¹Unter Dependenzgrammatik versteht man im Bereich des NLP die Untersuchung der hierarchischen Struktur von Sätzen und den Abhängigkeiten zwischen den Wörtern.

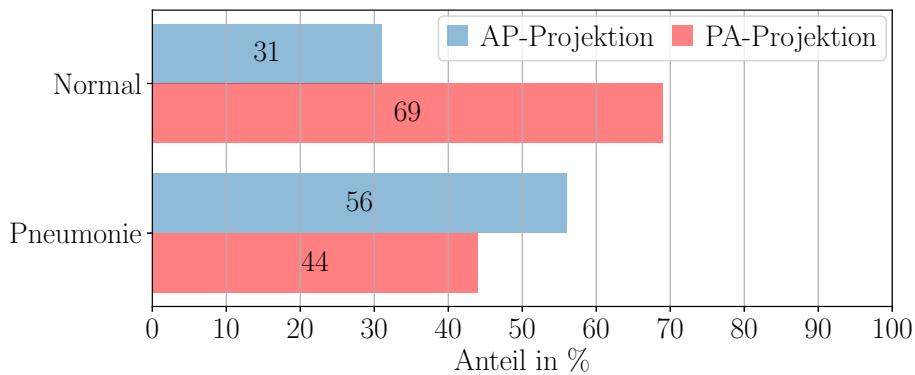


Abbildung 4.2: Bilder von Patienten mit Pneumonie sind im CXR14 Datensatz deutlich häufiger in AP-Projektion aufgenommen als bei gesunden Patienten

werden [vgl. 55, S. 4, Tabelle 2]. Vor der Verwendung dieses Datensatzes sollte somit unbedingt ein Bewusstsein für die Qualität der Label vorhanden sein.

Die Analyse der Metadaten ergab, dass das Geschlechterverhältnis mit 44 % weiblichen zu 56 % männlichen Patienten in etwa ausgeglichen ist. Die Anzahl der Samples mit dem Label *pneumonia* und *no finding* liegen bei 1 431, beziehungsweise 60 361 und weichen somit etwas von den Werten der Dokumentation ab [vgl. 84, S. 17, Tabelle 16]. Der Datensatz beinhaltet ausschließlich frontale Röntgenaufnahmen in PA- oder AP-Projektion, eine Filterung lateraler Ansichten war daher nicht notwendig. Das Verhältnis zwischen PA-Projektionen und AP-Projektionen ist deutlich davon abhängig, ob ein Patient als gesund eingestuft wurde, was in Abbildung 4.2 zu sehen ist. Diese Verteilung ist sehr kritisch, da eine Unterscheidung zwischen den Projektionen auf visueller Ebene problemlos möglich ist (siehe Abschnitt 3.4.2). Somit könnte das neuronale Netz während des Trainings fälschlicherweise lernen, dass Bilder in AP-Projektion automatisch auf eine Lungenentzündung hindeuten.

4.2.3 Pädiatrie Datensatz (Kermany)

Eine Forschungsgruppe um Daniel Kermany konnte 2017 ein populäres CNN erfolgreich auf zwei Klassifizierungsprobleme der medizinischen Bildgebung übertragen [37]. Für die Entwicklung des Modells machten sich die Forscher diverse Transfer Learning Techniken und vortrainierte Gewichte zunutze (siehe Abschnitt 5.1.2). Als Modellarchitektur wurde das aus der ImageNet Challenge bekannte Inception V3 Modell von Google verwendet [78]. Zum einen trainierten sie ein Modell für die Diagnostizierung von OCT-Aufnahmen und zum anderen eines für die Erkennung von Pneumonien in Thoraxröntgenbildern. Die erfolgreiche Anwendung der Architektur auf zwei völlig unterschiedliche Problemstellungen demonstriert, dass gute Ergebnisse einer Modellarchitektur auf andere Modalitäten und Anwendungsfälle übertragbar sind. Die Trainingsdaten für die Erkennung von Pneumonien stammen aus einem chinesischen Frauen- und Kinderkrankenhaus und wurden im Jahr 2018 für die Öffentlichkeit zugänglich gemacht [36]. Eine Übersicht der wichtigsten Eigenschaften

des Datensatzes können Tabelle 4.3 entnommen werden. Da der Datensatz durch die Autoren keinen Namen erhalten hat, wird er zur Vereinfachung im weiteren Verlauf dieser Arbeit als „Kermany Datensatz“ oder „Kermany“ referenziert.

Eigenschaft	Wert
Herausgeber	Daniel Kermany, Fang Zhang, Michael Goldbaum
Krankenhaus	Guangzhou Women and Children's Medical Center, China
Anzahl Patienten	5 856
Anzahl Röntgenbilder	5 856
Labeling-Methode	Manuell
Verfügbarkeit	Öffentlich zugänglich [36]

Tabelle 4.3: Übersicht des Pädiatrie Datensatz von Kermany u. a.

Die Bilder wurden von der pädiatrischen Abteilung des Krankenhauses zur Verfügung gestellt. Der medizinische Fachbegriff Pädiatrie bezeichnet hierbei das Teilgebiet der Kinder- und Jugendheilkunde. Die für den Datensatz ausgewählten Patienten befanden sich zum Zeitpunkt der Bildaufnahme im Alter zwischen ein und fünf Jahren. Alle Aufnahmen wurden im Rahmen von klinischen Routineuntersuchungen durchgeführt und befinden sich entweder in frontal AP Erect oder frontal AP Supine Projektion (siehe Tabelle 3.1, sowie Abbildungen 3.6d und 3.6e). Diese ist die Standardperspektive bei Kleinkindern, da für PA-Aufnahmen eine gewisse Körpergröße und die Fähigkeit ruhig und aufrecht zu stehen vorausgesetzt wird. Zu den einzelnen Bildern sind keine detaillierten Informationen wie Geschlecht oder Alter verfügbar. Eine Besonderheit der Daten im Vergleich zu den größeren Datensätzen ist die herausragende Qualität der Bilder und der Label. Ähnlich wie bei Open-i wurden alle Bilder einer manuellen Qualitätskontrolle unterzogen. Hierbei wurden alle qualitativ unzureichenden oder generell unleserlichen Aufnahmen entfernt. Die Bilder wurden anschließend von zwei Radiologen in die Klassen bakterielle Pneumonie, virale Pneumonie und normal eingeteilt. Die Bezeichnung normal bedeutet in diesem Kontext, dass keine Auffälligkeiten gefunden wurden und der Patient gesund zu sein scheint. Insgesamt beinhaltet der Datensatz 1 583 gesunde Patienten und 4 273 mit Pneumonie. Es herrscht somit ein deutliches Ungleichgewicht der Klassen vor, welches in Abschnitt 4.3.1 näher erläutert wird.

4.2.4 CheXpert

Zwei Jahre nach seiner Veröffentlichung wurde CXR14 (siehe Abschnitt 4.2.2) als größter öffentlich verfügbarer Datensatz für Thoraxröntgenbilder durch den etwa doppelt so großen CheXpert Datensatz abgelöst [28]. Er wurde von der Standford Machine Learning Group ins Leben gerufen, um die zuvor diskutierten Schwächen von CXR14 zu eliminieren und die Forschung im Bereich der CAD-Systeme voranzutreiben. Aufgrund der Anzahl und Qualität der Bilder, sowie die verbesserten Methodiken zur

Einteilung in Krankheitsklassen fand der Datensatz Zuspruch aus dem medizinischen Sektor [56]. Die Eckdaten von CheXpert sind in Tabelle 4.4 dargestellt.

Eigenschaft	Wert
Herausgeber	Stanford Machine Learning Group
Krankenhaus	Stanford Hospital, USA
Anzahl Patienten	65 240
Anzahl Röntgenbilder	224 316
Labeling-Methode	Regelbasiert
Verfügbarkeit	Öffentlich zugänglich [76]

Tabelle 4.4: Übersicht des CheXpert Datensatz

Ebenso wie bei CXR14 wurde ein NLP-Algorithmus zur Einteilung der Berichte in Krankheitsklassen entwickelt, welcher bei der Erstellung des Datensatzes besonders im Fokus stand. Es gibt jedoch einige nennenswerte Unterschiede zwischen den Labeling-Prozessen von CXR14 und CheXpert. Die Methode ist in drei aufeinanderfolgende Schritte aufgeteilt.

1. Extraktion von Krankheits-Erwähnungen
2. Unterscheidung zwischen positiven, negativen und unsicheren Erwähnungen
3. Gruppierung in 14 Klassen

Bei der Methode von Wang u. a. wird der erste und dritte Schritt direkt zu Beginn zusammen ausgeführt. Des Weiteren werden die Erwähnungen von Krankheiten bei CheXpert durch eine umfangreiche Sammlung von Phrasen und Regeln erkannt, welche händisch von mehreren zertifizierten Radiologen erstellt wurden [vgl. 28, S. 2]. Wie bereits in Abschnitt 4.2.2 erwähnt, erfolgt dieser Schritt bei CXR14 automatisiert durch MetaMap und DNorm. Zur Qualitätssicherung des Systems wurde ein Evaluationsdatensatz mit 1 000 Radiologieberichten erstellt, welche von zwei zertifizierten Radiologen unabhängig voneinander beurteilt wurden. Die Autoren geben zudem an, im direkten Vergleich mit dem in CXR14 verwendeten NegBio-Labeler auf den Evaluationsdaten in der Extraktion von Erwähnungen, sowie der Erkennung von Negationen und Unsicherheiten signifikant besser abzuschneiden [vgl. 28, S. 3].

Die Verfügbaren Label unterteilen sich in 12 Krankheiten, sowie die Sonderfälle *No Finding* und *Support Devices*. Unter einem *Support Device* versteht man medizinische Hilfsgeräte wie Herzschrittmacher oder Thoraxdrainagen, die auf Röntgenbildern zu sehen sind. Das Label *No Finding* wird dann vergeben, wenn keine der 12 Krankheiten ein positives oder unsicheres Label erhielten. Medizinische Hilfsgeräte sind eher bei kranken, als bei gesunden Patienten zu erwarten, da diese zur Behandlung von Krankheiten benötigt werden. Wie in Abbildung 4.3 zu sehen ist konnte diese Annahme für die vorliegenden Daten jedoch nicht bestätigt werden. Der Anteil gesunder Patienten mit medizinischen Hilfsgeräten ist um knapp 10 % höher als bei Patienten mit Pneumonie.

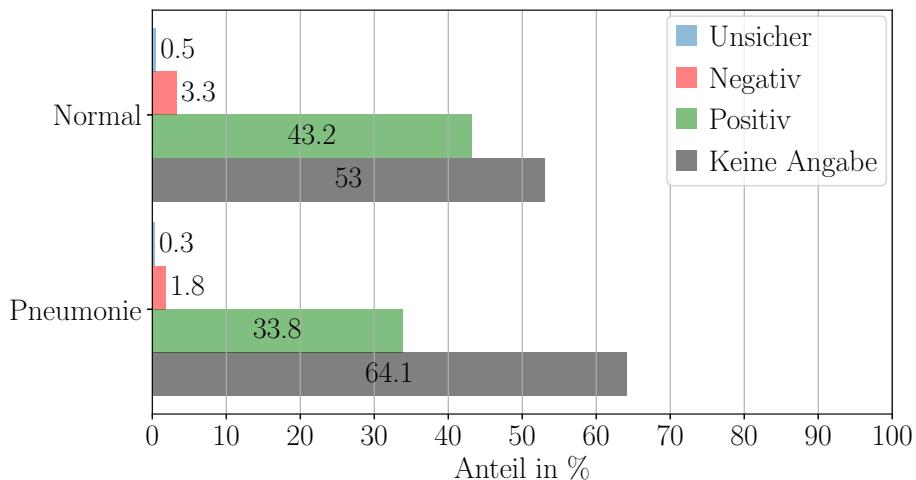


Abbildung 4.3: Die Verteilung der medizinischen Hilfsgeräte im CheXpert Datensatz zeigt, dass das Label *Support Devices* bei gesunden Patienten häufiger vergeben wurde als bei Patienten mit Pneumonie

Im Umgang mit Negationen und Unsicherheiten konnten ebenfalls Fortschritte erreicht werden. Anstelle aller unsicheren und negierten Erwähnungen aus den Metadaten zu entfernen, werden diese mit in den Datensatz integriert, da diese Informationen bei der Entwicklung von CAD-Systemen relevant sein könnten. Für jedes Label können daher die Werte 1 (positiv), 0 (negativ), -1 (unsicher) und nan (keine Erwähnung) vergeben werden. Eine Unsicherheit impliziert, dass es nicht eindeutig möglich war, eine Entscheidung für oder gegen eine Krankheitsklasse zu fällen. Die Autoren von CheXpert schlagen drei triviale, sowie zwei komplexe Konzepte für den Umgang mit unsicheren Labels vor [vgl. 28, S. 3f]. Um die Konzepte zu evaluieren, wurde für jeden Ansatz ein neuronales Netz zur Klassifizierung von Röntgenbildern trainiert. Der einfachste Umgang mit unsicheren Labels besteht darin, entweder eine Substitution mit positiven (U-Ones), beziehungsweise negativen (U-Zeros) Labels durchzuführen, oder unsichere Labels generell zu verwerfen (U-Ignore). Als eine der komplexen Herangehensweisen wird vorgeschlagen, die binäre Klassifikation in eine ternäre Aufgabe zu überführen, indem das Label für Unsicherheit als dritte Klasse der jeweiligen Zielvariable eingeführt wird (U-MultiClass). Abschließend schlagen Irvin u. a. vor, im ersten Schritt ein neuronales Netz zur Klassifikation von Röntgenbildern mit dem U-Ignore Ansatz zu trainieren. Dieses Netz soll anschließend verwendet werden, um die Klasse der unsicher gelabelten Samples anhand ihrer Röntgenbilder vorherzusagen. Die vorhergesagten Klassen werden dann anstelle der unsicheren Label in den Datensatz eingetragen (U-SelfTrained). Da sich bei der Evaluation durch die Autoren keine der fünf Herangehensweisen besonders abheben konnte, wurde für diese Arbeit der naheliegende U-Ignore Ansatz gewählt. Der U-SelfTrained Ansatz ist der Einzige, der bei allen fünf evaluierten Krankheitsklassen besser als die U-Ignore Methode abschneidet, wenn auch nur geringfügig. Er wurde daher ebenfalls in Betracht gezogen und könnte im Rahmen einer weiteren Arbeit verwendet werden, um die Menge verfügbarer Daten zu erweitern. Die Anzahl der unsicheren Labels für Pneumonien liegt laut Dokumentation bei 15 658 und bietet damit ein nicht zu vernachlässigendes

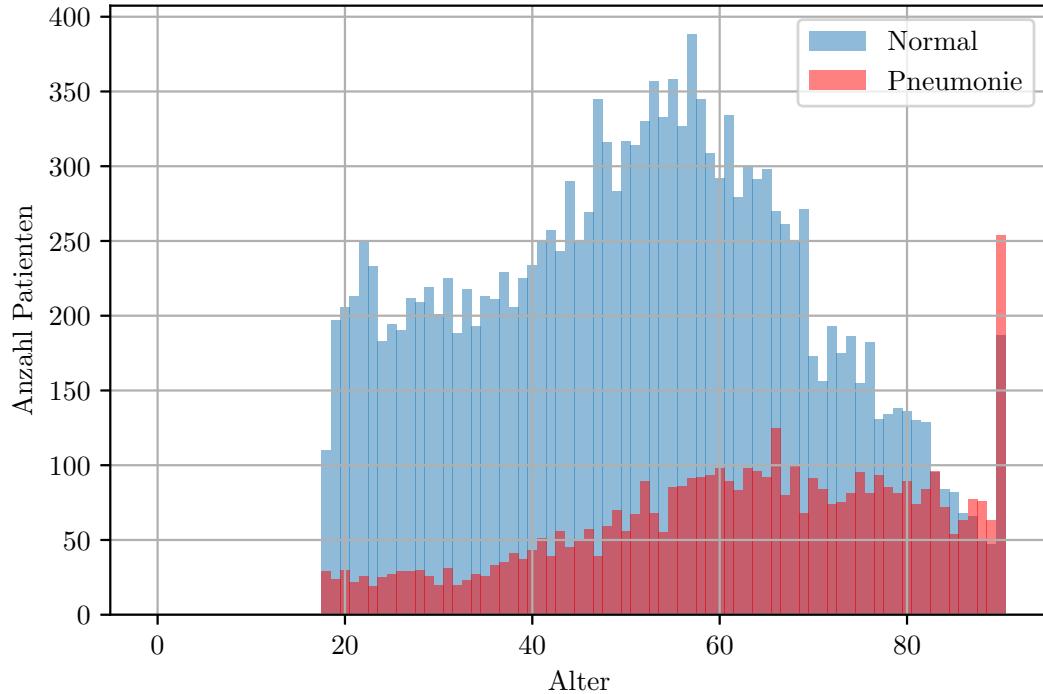


Abbildung 4.4: Übersicht der Altersverteilung von Patienten mit und ohne Pneumonie im CheXpert Datensatz in Form eines Histogramms

Potential die Trainingsdaten zu erweitern [vgl. 28, S. 2].

Zusätzlich zu den Krankheitsklassen der Radiologieberichte stehen für jedes Röntgenbild Informationen über Alter, Geschlecht, Projektion (PA oder AP) und Seitigkeit (frontal oder lateral) zur Verfügung. Da laterale Bilder häufig nur als Zusatzinformation für frontale Röntgenbilder verwendet werden, verringerte sich die Gesamtzahl der Patienten durch das Entfernen der 32 419 lateralen Bilder lediglich um 6 Patienten. Der Datensatz enthält nach dem Filtern lateraler Ansichten noch 4 683 Bilder mit positivem Pneumonie Label und 17 000 Bilder, bei denen keine Krankheit festgestellt wurde. Die Anzahl der Patienten liegt bei 3 935 (Pneumonie) und 14 027 (Normal). Somit ist das häufig kritisierte Verhältnis zwischen der Anzahl der Patienten und der Anzahl der Röntgenbilder nach der Filterung erheblich besser als im gesamten CheXpert Datensatz ist [56]. Für die verbliebenen Samples wurden anschließend die Verteilungen von Projektion, Geschlecht und Alter untersucht. Im Vergleich zu CXR14 (siehe Abbildung 4.2) ist der Anteil von AP-Projektionen bei Patienten mit (74,4 %) und ohne Pneumonie (67,3 %) deutlich ähnlicher. Die Geschlechterverteilung liegt bei etwa 42 % weiblichen und 58 % männlichen Patienten, wobei auch hier die Differenzen zwischen Patienten mit und ohne Pneumonie nur sehr gering sind ($\pm 1\%$). Die Vereinigungsnorm aufgrund von Projektion oder Geschlecht ist in diesem Fall somit nur marginal. Die Untersuchung der Altersverteilung zeigt, dass alle Patienten mindestens 18 Jahre alt sind, was vermutlich auch rechtliche Gründe zurückzuführen ist. Das Histogramm in Abbildung 4.4 zeigt, dass im höheren Altersbereich ab etwa 60 Jahren

große Unterschiede zwischen den Klassen zu erkennen sind. Dort nimmt die Anzahl der gesunden Patienten stark ab, während die Anzahl von Patienten mit Pneumonie größtenteils konstant bleibt. Eine bevölkerungsbasierte Überwachungsstudie aus Ohio in den USA kam zu dem Ergebnis, dass die Inzidenz von Pneumonien bei Erwachsenen mit steigendem Alter zunimmt [44]. Somit spiegelt die altersabhängige Verteilung der Pneumoniefälle im Datensatz die tatsächliche Verteilung in der Weltbevölkerung wider. Die außergewöhnlich hohe Anzahl Patienten im Alter von exakt 90 Jahren ist durch die Dokumentation des Datensatzes nicht begründet. Aufgrund des deutlichen Bias in der Altersverteilung ist es enorm wichtig diese Eigenschaft bei der Stratifikation in Abschnitt 4.3.2 zu berücksichtigen.

4.2.5 MIMIC-CXR

Nur kurze Zeit nach CheXpert wurde im gleichen Jahr der MIMIC-CXR Datensatz veröffentlicht. Er ist der bis dato größte öffentlich verfügbare Datensatz für Thoraxröntgenaufnahmen und wurde unter anderem durch das Unternehmen Philips Healthcare finanziert [vgl. 32, S. 6]. Hauptsächlich für die Entwicklung verantwortlich war das Massachusetts Institute of Technology (MIT) in Zusammenarbeit mit den Institutionen Beth Israel Deaconess Medical Center, Stanford University, Harvard Medical School und NIH. Der Datensatz wurde im Digital Imaging and Communications in Medicine (DICOM)-Format und im Joint Photographic Experts Group (JPEG)-Format separat veröffentlicht [33][32]. Wenn in dieser Arbeit der Begriff MIMIC-CXR verwendet wird, ist dabei stets die JPEG-Variante des Datensatzes gemeint.

Eigenschaft	Wert
Herausgeber	MIT Laboratory for Computational Physiology
Krankenhaus	Beth Israel Deaconess Medical Center, USA
Anzahl Patienten	65 379
Anzahl Röntgenbilder	377 110
Labeling-Methode	Regelbasiert
Verfügbarkeit	Öffentlich zugänglich, Zertifizierungs- und Verifizierungsverfahren notwendig [31]

Tabelle 4.5: Übersicht des MIMIC-CXR Datensatz

Ebenso wie bei den anderen großen Datensätzen wurden die Auffälligkeiten der Röntgenbilder anhand der Radiologieberichte abgeleitet. Für die grundsätzliche Verarbeitung der Berichte wurde die gleiche Software und die gleiche Vorgehensweise wie bei CheXpert verwendet (siehe Abschnitt 4.2.4 und [28]). Lediglich bei der Unterscheidung zwischen positiven, negativen und unsicheren Erwähnungen wurde separat zum CheXpert-Labeler der NegBio-Labeler angewandt, welcher bereits bei CXR14 verwendet wurde [60]. Die Ergebnisse beider Varianten sind unabhängig voneinander im Datensatz enthalten, um dem Nutzer die Entscheidung zu überlassen, welche Label er für die Röntgenbilder verwenden möchte. Da der NegBio-Labeler während der

Evaluation für die Pneumoniekasse minimal besser abgeschnitten hat, wurden in dieser Arbeit die NegBio-Label verwendet [vgl. 32, S. 5f, Tabelle 4-6]. Weiterhin zeigen Johnson u. a., dass die beiden bezug auf die Klassifizierung von Pneumonien sich bei 99,4 % der 227 827 Radiologieberichte einig waren [vgl. 32, S. 4, Tabelle 2].

Die Metadaten enthalten keine persönlichen Informationen über die Patienten. Daher konnten bei der Untersuchung der Daten lediglich die Projektion der Bilder und die Präsenz medizinischer Hilfsgeräte analysiert werden. Auch hier sollten diese Informationen genutzt werden, um seitliche Röntgenaufnahmen zu entfernen und einen möglichen Bias zwischen Patienten mit und ohne Pneumonie zu identifizieren. Das Feld *ViewPosition* beinhaltet Informationen über die verwendete Projektion. Bei 15 769 von 377 110 Bildern ist dieses Feld jedoch leer, was einem nicht vernachlässigbaren Anteil von 4,18 % des gesamten Datensatzes entspricht. Um die fehlende Information zu erschließen, wurden die weiteren Felder dieser Samples untersucht. Die Felder *ProcedureCodeSequence_CodeMeaning* und *PerformedProcedureStepDescription* beinhalten Informationen über das bei der jeweiligen Studie eingesetzte Diagnoseverfahren, welches Aufschluss über die verwendete Projektion bieten kann. Um die Untersuchung beider Felder zu vereinfachen, wurden deren Inhalte verkettet und in dem neuen Feld *ProcedureCombined* im dataframe gespeichert. Die Analyse ergab, dass in 85,4 % der Fälle das Diagnoseverfahren *CHEST (PA AND LAT)* und in 14,4 % der Fälle *CHEST (PORTABLE AP)* angewandt wurden. Die verbleibenden 0,2 % setzen sich aus einer Vielzahl vernachlässigbarer Kategorien mit jeweils weniger als 10 Bildern zusammen und wurden daher nicht berücksichtigt.

Bei Samples mit dem Label *CHEST (PA AND LAT)* handelt es sich um Studien mit mindestens zwei Röntgenaufnahmen, bei denen sowohl die PA-Projektion als auch die L-Projektion zum Einsatz kam. Da sich keine zuverlässige Systematik in den restlichen Informationen wie Dateinamen, Bildauflösung oder zeitliche Reihenfolge erkennen ließ, konnte nicht erschlossen werden welche Projektion bei der jeweiligen Aufnahme verwendet wurde. Um sicherzustellen, dass alle seitlichen Röntgenaufnahmen entfernt werden, konnten diese Samples daher nicht verwendet werden.

Samples mit dem Label *CHEST (PORTABLE AP)* beinhalten Aufnahmen in AP-Projektion, die durch ein mobiles Röntgengerät entstanden sind. Da sich die Bilder mobiler Röntgengeräte von Bildern stationärer Röntgengeräte unterscheiden können, ist diese Information für die Untersuchung des Bias relevant. Daher wurde für diese Bilder ein neues, nicht vom Datensatz zur Verfügung gestelltes Label *AP_horizontal* im Feld *ViewPosition* eingetragen. Durch die Analyse des Diagnoseverfahrens konnten somit für 2 252 von 15 769 Samples die fehlende Information über die Projektion erschlossen und zusätzliche Informationen über die Röntgenbilder offenbart werden. Nach der Bereinigung seitlicher Bilder befanden sich noch 16 807 Samples mit Pneumonie, sowie 66 959 gesunde Samples im Datensatz.

Im Vergleich zwischen Patienten mit und ohne Pneumonie ist drastischer Unterschied bei der Verwendung von mobilen Röntgengeräten zu erkennen (siehe Abbildung 4.5). Innerhalb dieses Datensatzes besteht somit ein unmittelbarer Zusammenhang zwischen der Nutzung mobiler Röntgengeräte und dem Gesundheitszustand des Patienten. Diese Erkenntnis kann aus medizinischer Sicht dadurch hinterlegt werden, dass

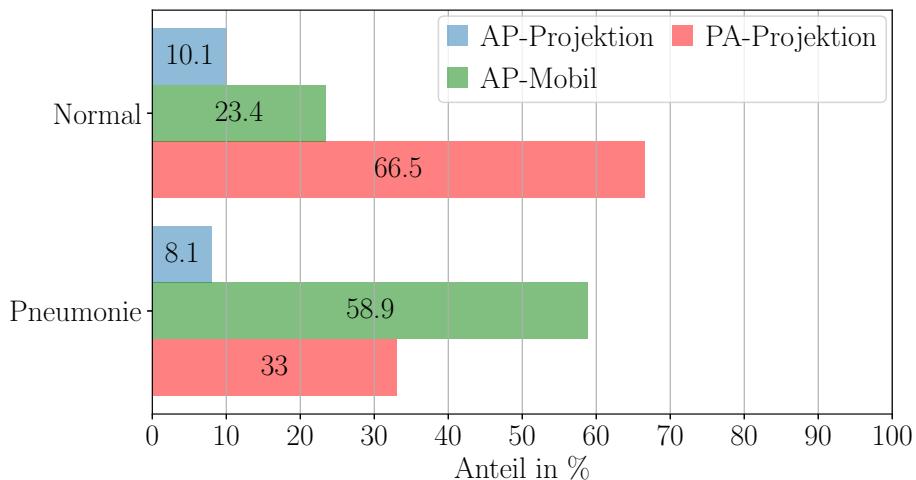


Abbildung 4.5: Verteilung der Projektionsart bei Samples mit und ohne Pneumonie im MIMIC-CXR Datensatz

besonders schwer erkrankte Patienten häufig nicht transportfähig sind. Die Mediziner Rubinowitz u. a. geben an, dass im Yale-New Haven Hospital etwa die Hälfte aller Thoraxröntgenbilder durch portable Geräte auf der Intensivstation entstehen [vgl. 73, S. 540]. Weiterhin erklärt der Artikel, dass Röntgenaufnahmen für die Diagnose vieler Krankheiten unverzichtbar sind und die geringere Qualität im Vergleich zu Aufnahmen stationärer Röntgengeräte daher toleriert werden muss. Durch die Unterschiede der Bilder ist somit anzunehmen, dass ein neuronales Netz in der Lage ist, mobile Röntgengeräte anhand der Röntgenbilder zu erkennen und diese Information in Kombination mit dem vorliegenden Bias sich negativ auf die Vorhersagen des Modells auswirkt. Unabhängig von der Art des Röntgengeräts konnte zudem festgestellt werden, dass der Anteil Röntgenbilder in AP-Projektion höher bei Patienten ist, die an einer Pneumonie erkrankt sind. Diese Korrelation konnte zuvor bereits bei CXR14 in Abbildung 4.2 beobachtet werden. Der alarmierende Bias bei der Verwendung mobiler Röntgengeräte und der eingesetzten Projektion muss bei der nachfolgenden Stratifikation der Daten in Abschnitt 4.3.2 dringend beachtet werden.

Im Gegensatz zum CheXpert Datensatz bestätigte sich bei MIMIC-CXR die Annahme, dass bei kranken Patienten tendenziell häufiger medizinische Hilfsgeräte auf den Röntgenbildern zu sehen sind. In Abbildung 4.6 ist zu sehen, dass Patienten mit Pneumonie sogar fast drei Mal häufiger medizinische Hilfsgeräte tragen als gesunde Patienten. Dieser Zusammenhang sollte äußerst kritisch hinterfragt werden, da die Geräte auf den Röntgenbildern sehr leicht zu erkennen sind und somit die Gefahr besteht, dass dadurch die das Vorhersagemodell beeinflusst wird. Ergänzend ist anzumerken, dass in diesem Datensatz generell weit weniger positive Label im Feld *Support Device* vergeben wurden als bei CheXpert.

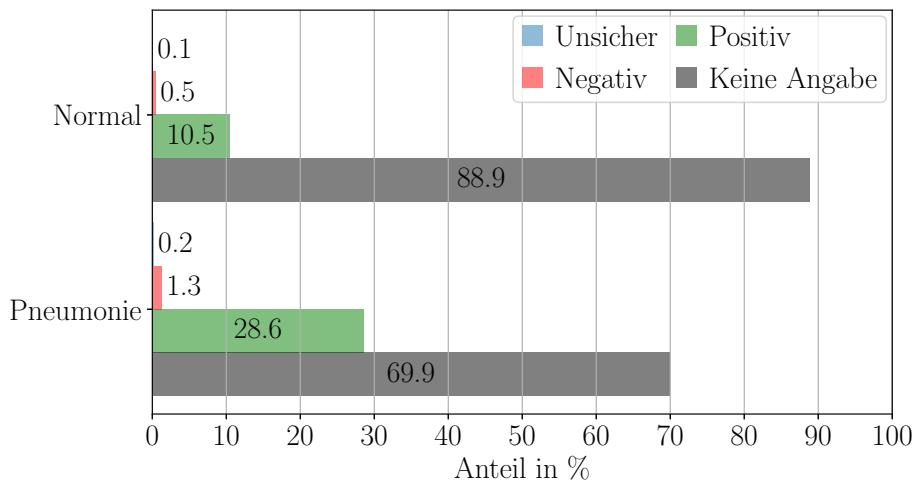


Abbildung 4.6: Verteilung der medizinischen Hilfsgeräte bei Patienten mit und ohne Pneumonie im MIMIC-CXR Datensatz

4.2.6 PadChest

PadChest erschien im Jahr 2019 und ist zum Zeitpunkt dieser Arbeit der aktuellste öffentliche Datensatz für Thoraxröntgenaufnahmen [10]. Bei der Entwicklung des Datensatzes wurden teils andere Methodiken verwendet, wodurch er sich in vielerlei Hinsicht von den bisher diskutierten Datensätzen abhebt. Die Röntgenberichte und -bilder stammen aus einem Krankenhaus in Spanien, was ihn zum ersten Datensatz dieser Größenordnung macht, der nicht auf Radiologieberichten in englischer Sprache basiert. Abgesehen von den sprachlichen Unterschieden ist PadChest der erste große Datensatz, der aus Europa stammt.

Eigenschaft	Wert
Herausgeber	Banco digital de Imagen Medica de la Comunidad Valenciana (BIMCV)
Krankenhaus	Hospital San Juan de Alicante, Spanien
Anzahl Patienten	67 625
Anzahl Röntgenbilder	160 868
Labeling-Methode	Deep Learning-basiert
Verfügbarkeit	Öffentlich zugänglich [6]

Tabelle 4.6: Übersicht des PadChest Datensatz

Die Vorgehensweise und Genauigkeit beim Labeling der Röntgenberichte ist für die Qualität des Datensatzes von zentraler Bedeutung. Da die Berichte in Spanisch verfasst sind und die bestehenden Lösungen wie NegBio oder der CheXpert-Labeler auf die englische Sprache zugeschnitten sind, musste ein neues Verfahren entwickelt werden. Anstatt die bekannten regelbasierten Systeme für den neuen Anwendungsfall anzupassen, entschieden sich Bustos u. a. dazu, einen neuen Weg einzuschlagen

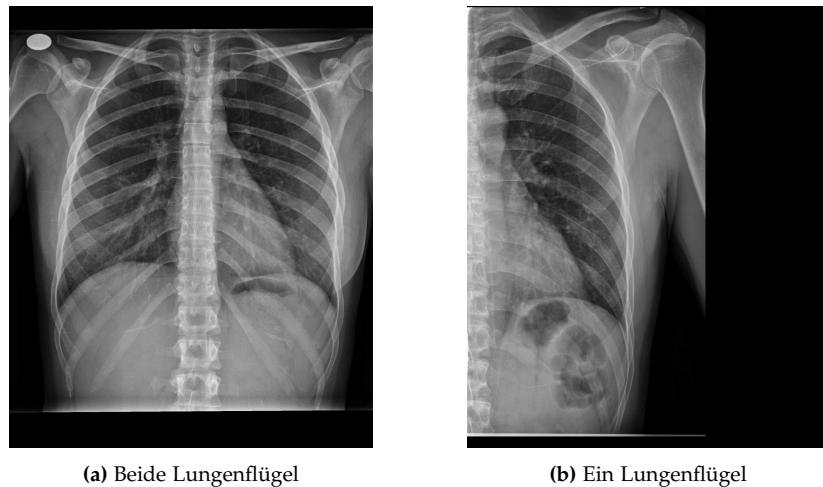


Abbildung 4.7: Vergleich von zwei Röntgenbildern mit der Projektion COSTAL (Quelle: [6])

und ein Deep Learning-basiertes System für die Verarbeitung der Röntgenberichte zu entwickeln. Zur Vorverarbeitung wurde das Sprachverarbeitungsprogramm NLTK [43] verwendet, um Stopwörter zu entfernen, Wortstämme zu erkennen und die Texte in Sätze zu zerlegen. Insgesamt bestanden die Berichte aus etwa einer halben Millionen Sätzen. Durch die Anwendung von Doc2Vec [41] als Clustering-Methode wurden die Sätze in 20 Themenbereiche unterteilt. Diese Einteilung erleichterte den Medizinern die Beurteilung der Sätze und steigerte somit die Effizienz des manuellen Labeling-Prozesses. Insgesamt wurden 27 593 Radiologieberichte von geschultem Fachpersonal gelabelt, was 27 % der gesamten Berichte entspricht. Im Vergleich zu den bisher diskutierten Datensätzen beinhaltet PadChest wesentlich mehr Labels, welche sich aus 19 Differentialdiagnosen, 174 radiologischen Auffälligkeiten und 104 anatomischen Regionen zusammensetzen. Die Labels wurden anschließend in drei separate hierarchische Baumstrukturen unterteilt, deren Aufbau in der Originaldokumentation des Datensatzes zu finden ist [vgl. 10, S. 26-32]. Durch die insgesamt 297 Labels ist der Informationsgehalt dieses Datensatzes enorm umfangreich, was die Entwicklung von komplexeren CAD-Systemen begünstigt. Auf Basis der manuell beurteilten Radiologieberichte konnte ein Recurrent Neural Network (RNN) mit sogenanntem Attention Mechanism trainiert werden, welches zum Labeln der verbliebenen Berichte eingesetzt wurde. Zur Qualitätskontrolle wurde das Modell auf 500 manuell gelabelten Sätzen evaluiert, welche nicht Teil der Trainingsdaten waren.

Ebenso wie bei den vorherigen Datensätzen wurden laterale Bilder aus dem Datensatz entfernt und Bilder extrahiert, die entweder keine Krankheiten vorweisen, oder als Pneumonie klassifiziert wurden. Weiterhin beinhaltet PadChest Samples, deren Projektion als COSTAL gekennzeichnet sind. Diese Aufnahmen sind darauf ausgelegt, dass Knochen besonders gut sichtbar sind und werden vor allem verwendet, wenn Verletzungen an den Rippen untersucht werden sollen. Wie in Abbildung 4.7b zu sehen ist, wird bei einseitigen Verletzungen der Rippen eventuell nur eine Hälfte des Oberkörpers untersucht. Diese abgeschnittenen Röntgenaufnahmen sind für die Klassifikation von Pneumonien ungeeignet, da sie nur einen Lungenflügel darstellen. Aufgrund der vernachlässigbaren Anzahl wurden Bilder mit der Projektion COSTAL

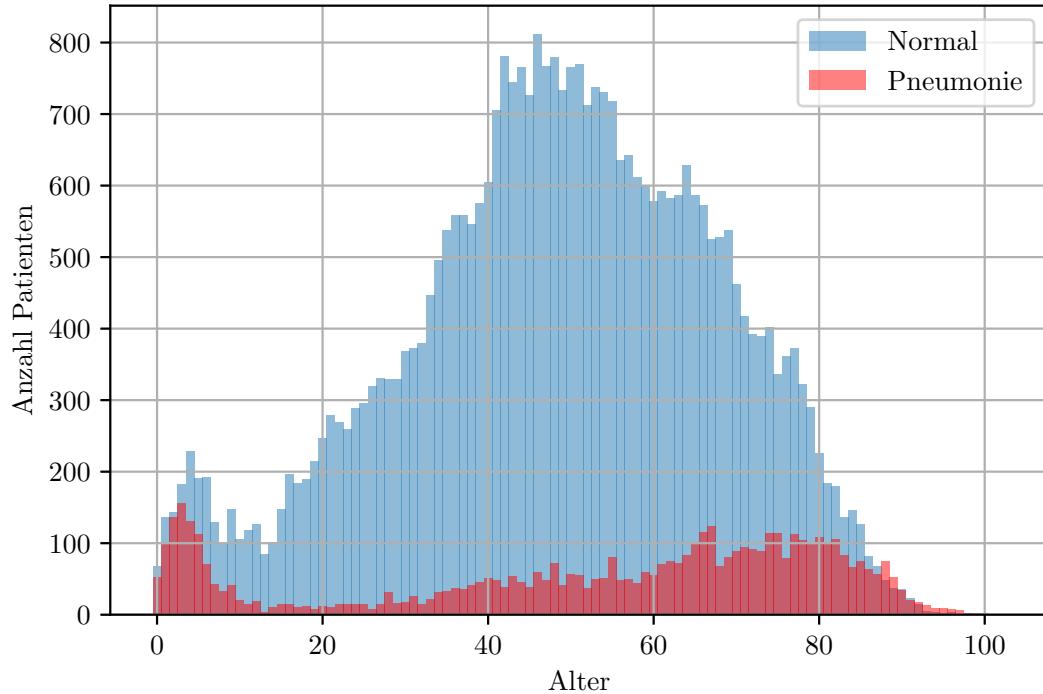


Abbildung 4.8: Verteilung des Alters der Patienten im PadChest Datensatz

vorsorglich aus dem Datensatz entfernt. Nach der Bereinigung anhand von Projektion und Krankheitslabel beinhaltet der Datensatz noch 5 220 Bilder mit Pneumonien, sowie 35 915 gesunde Samples. Die wichtigsten Eigenschaften dieser Samples werden nachfolgend gegenübergestellt.

Wie in Abbildung 4.8 gut zu erkennen ist, unterscheidet sich die Altersverteilung der gesunden Patienten deutlich von der Verteilung der Patienten mit Pneumonie. Vor allem Kinder im Alter von 0 bis 8 Jahren und sehr alte Patienten ab 80 Jahren, sind häufig von Lungenentzündungen betroffen. Im Bereich von 20 bis 80 Jahren ist bei den als normal gelabelten Patienten annähernd eine Gauß-Verteilung zu beobachten. Im Gegensatz dazu steigt die Anzahl der Patienten mit Pneumonie in diesem Altersbereich kontinuierlich an. Die hier beobachteten Altersverteilungen für Patienten mit und ohne Pneumonie decken sich dabei mit den Verteilungen im CheXpert Datensatz (siehe Abbildung 4.4). Die besonders hohe Anzahl an Kindern mit Pneumonie ist dabei durch Studien zu erklären. Ein Bericht der WHO macht darauf aufmerksam, dass Pneumonien die häufigste Todesursache bei Kindern bis fünf Jahren darstellen [9].

In den Daten wird zwischen PA-Erect, AP-Erect (siehe Abbildung 3.6d) und AP-Horizontal Projektion (siehe Abbildung 3.6e) unterschieden. Die Verteilung der Projektionsarten im bereinigten PadChest Datensatz ist in Abbildung 4.9a zu sehen. Aufgrund der schlechteren Bildqualität wird AP-Horizontal in der Regel nur verwendet, wenn eine aufrechte Lagerung des Patienten aus gesundheitlichen oder anderen Gründen nicht möglich ist. Allgemein ist zu erkennen, dass bei Patienten mit Pneumonie die

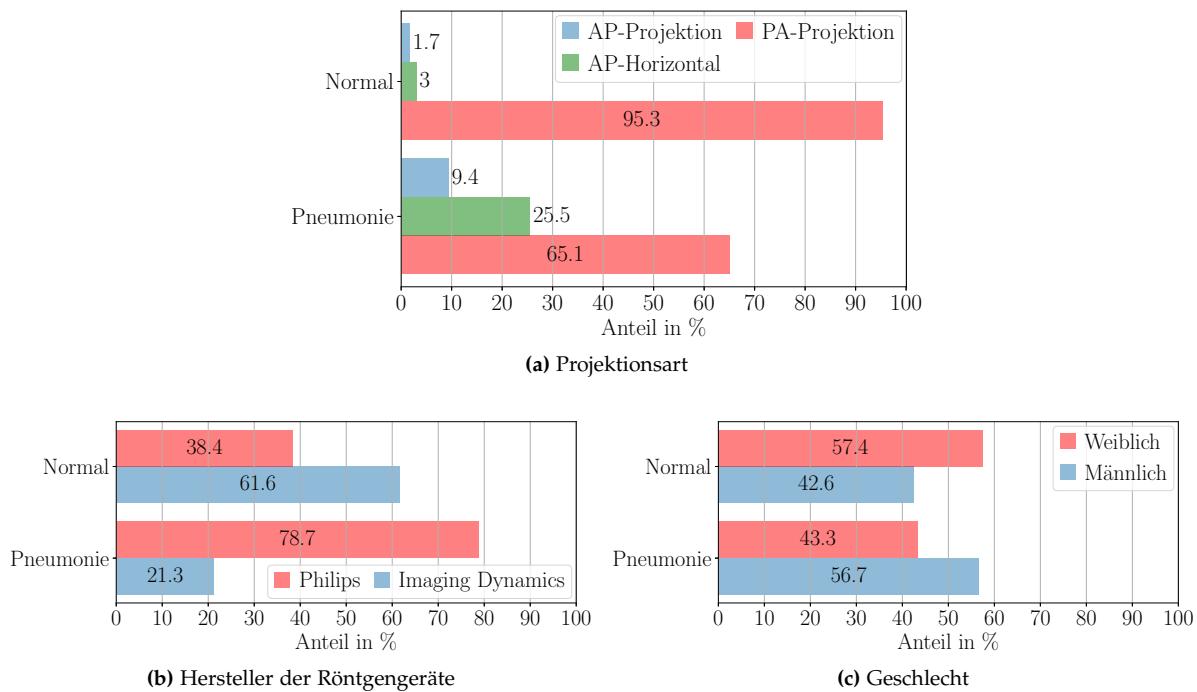


Abbildung 4.9: Vergleich der Verteilung verschiedener Eigenschaften bei Samples mit und ohne Pneumonie im PadChest Datensatz

PA-Projektion seltener eingesetzt wird. Dieser Zusammenhang ist in unterschiedlichem Ausmaß in allen großen Datensätzen vorzufinden (CXR14, CheXpert, MIMIC-CXR, PadChest). Weiterhin ist in Abbildung 4.9b zu sehen, dass deutliche Unterschiede abhängig vom Hersteller des Röntgengeräts präsent sind. Eine mögliche Begründung für die unterschiedlich hohen Anteile ist der Einsatzort des jeweiligen Gerätes. So ist es naheliegend, dass im Hospital San Juan de Alicante Geräte der Firma Philips aufgrund besonderer Eigenschaften vermehrt auf den Intensivstationen verwendet werden. Routineuntersuchungen überwiegend gesunder Patienten würden somit häufiger mit Geräten des Herstellers Imaging Dynamics durchgeführt werden. Besonders auffällig ist, dass alle Bilder mit der Projektion AP-Horizontal von Röntgengeräten der Firma Philips aufgenommen wurden. Diese Feststellung bekräftigt die zuvor formulierte Vermutung des spezifischen Einsatzorts der Geräte. Die Überrepräsentation bestimmter Eigenschaften aufgrund des Einsatzortes ähnelt den Beobachtungen bezüglich mobiler Röntgengeräte bei MIMIC-CXR (Abbildung 4.5). Selbstverständlich ist auch in diesem Fall der Bias hochproblematisch. Der Hersteller des Gerätes könnte beispielsweise indirekt durch herstellerspezifische Markierungen oder weiterer Eigenschaften der Bilder identifizierbar sein. Für einen Arzt sind diese Eigenschaften während der Befundung zwar irrelevant, jedoch könnte der damit verbundene Bias dafür sorgen, dass ein neuronales Netz lernt, abhängig vom Hersteller des Röntgengeräts eine Krankheit zu detektieren.

Die zuvor bereits erwähnte bevölkerungsisierte Studie von Marston u. a. zeigt nicht nur, dass Pneumonien mit steigendem Alter häufiger auftreten, sondern auch, dass Männer etwas häufiger erkranken als Frauen [vgl. 44, S. 1712]. Wie in Abbildung 4.9c zu

sehen ist, konnte sich diese Korrelation während der Analyse des PadChest Datensatz bestätigen. In den zuvor diskutierten Datensätzen war dieser Zusammenhang nicht erkennbar.

4.3 Auswahl der Daten

Der vorherige Abschnitt beschreibt die Entstehung der Datensätze, Bereinigung unerwünschter Röntgenbilder und analysiert die Verteilung der Eigenschaften. Die Anzahl der gesunden Patienten übersteigt die Anzahl der Patienten mit Pneumonie dabei deutlich, weshalb Abschnitt 4.3.1 eine Strategie zum Umgang mit diesem Missverhältnis entwickelt. Anschließend daran entwickelt Abschnitt 4.3.2 einen Algorithmus zur Reduktion des bereits ausführlich diskutierten Bias der Datensätze. Die beiden letzten Abschnitte kombinieren die entwickelten Datensätze miteinander und beschreiben die Zusammensetzung des geschaffenen Gesamtdatensatz.

4.3.1 Balancierung der Klassen

Gesunde Patienten sind im Vergleich zu Patienten mit Pneumonie in den Datensätzen stark überrepräsentiert, was Abschnitt 4.2 gezeigt hat. Dieses Ungleichgewicht sorgt dafür, dass die Verlustfunktion die unterrepräsentierte Klasse weniger berücksichtigt. Um dieser Situation entgegenzuwirken kann auf die Klassen eine Gewichtung, auf Grundlage der Häufigkeit im Datensatz angewendet werden [vgl. 14, S. 3f]. Diese Vorgehensweise bietet sich besonders an, wenn die Menge der Trainingsdaten begrenzt ist und das Verwerfen von Daten keine Option darstellt. Die Gewichtung der Klassen und Beibehaltung aller Daten bringt jedoch den Nachteil mit sich, dass die Missverhältnisse in der Verteilung der Eigenschaften erhalten bleiben. Die Auswirkung der Klassenzugehörigkeit auf den Trainingsprozess ist somit zwar ausgeglichen, aber der Bias der Eigenschaften bleibt unverändert.

Die zuvor analysierten Datensätze beinhalten insgesamt etwa eine Millionen Thoraxröntgenbilder. Zum Ausbalancieren der Klassen kann es sich diese Arbeit daher erlauben, Bilder von gesunden Patienten zu verwerfen. Die Anzahl verbleibender gesunder Patienten richtet sich nach der Anzahl der Patienten mit Pneumonie, um das Gleichgewicht zwischen den Klassen herzustellen. Diese Vorgehensweise geht durch das Verwerfen von Röntgenbildern offensichtlich mit einem Informationsverlust einher. Sie bietet jedoch die besondere Gelegenheit, eine Auswahl und Modellierung der Daten anhand ihrer Eigenschaften durchzuführen, die zum Vorteil des Modelltrainings angepasst werden kann. Dieser Schritt ist besonders wichtig, da sich die ungleichen Verteilungen des gesamten Datensatzes sonst im kleinen Maßstab auch nach dem Ausbalancieren der Klassen in den Daten widerspiegeln. Dieser Zusammenhang ist in Abbildung 4.10 am Beispiel der Altersverteilung im PadChest Datensatz dargestellt.

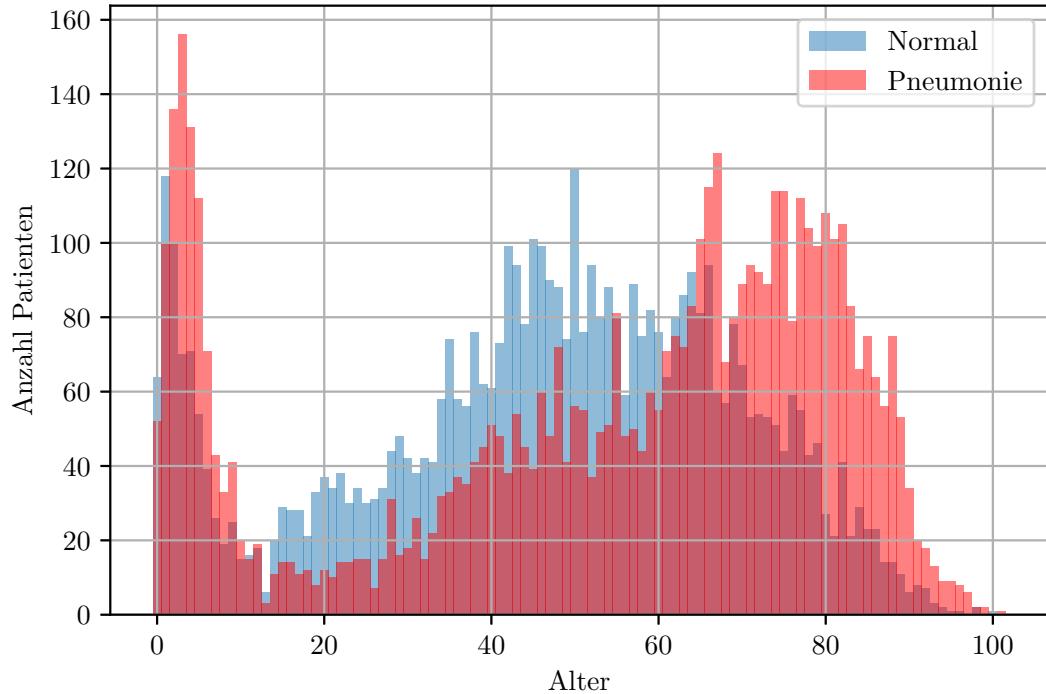


Abbildung 4.10: Die unterschiedliche Altersverteilung zwischen Patienten mit und ohne Pneumonie im PadChest Datensatz wird beim Ausbalancieren der Klassen übernommen, wenn keine Stratifikation angewendet wird

Das Histogramm zeigt, dass sich bei einer zufälligen Auswahl der gesunden Patienten die Altersverteilungen voneinander unterscheiden, was bereits für den gesamten Datensatz in Abbildung 4.8 beobachtet werden konnte.

4.3.2 Stratifikation

Der vorherige Abschnitt hat bereits angedeutet, dass durch selektives Verwerfen von Daten eine Modellierung ihrer Eigenschaften erreicht werden kann. Um dem Ungleichgewicht der Klassen und ihrer Eigenschaften entgegenzuwirken, wird nachfolgend eine Methode entwickelt, die von der stratifizierten Zufallsstichprobe inspiriert ist. Die stratifizierte Zufallsstichprobe, auch geschichtete Zufallsstichprobe genannt, ist eine Vorgehensweise aus der Statistik, bei der Teilmengen einer Population ausgewählt werden. Der Grundgedanke der Methode besteht darin, die proportionale Verteilung der Merkmale der Population beim Ziehen der Stichprobe zu berücksichtigen, sodass sich diese Merkmale auch in der Teilmenge entsprechend wiederfinden [vgl. 12, S. 89].

Die hier dargelegte Problemstellung weicht etwas von der Ausgangssituation der stratifizierten Zufallsstichprobe ab. Es wird zwar eine Teilmenge der gesunden Patienten ausgewählt, allerdings sollen deren Merkmale sich nicht an der Grundgesamtheit

der gesunden, sondern der Grundgesamtheit der kranken Patienten orientieren. Die Implementierung erfasst daher zuerst die proportionalen Verteilungen der Pneumoniedaten und projiziert diese Eigenschaften anschließend bei der Datenauswahl auf die Daten der gesunden Patienten. Um die zuvor beschriebene Balancierung der Daten durchzuführen, wird die Gesamtgröße der Teilmenge so gewählt, dass die Anzahl der Patienten mit und ohne Pneumonie nach der Stratifikation identisch sind.

Die Umsetzung des beschriebenen Konzepts zur stratifizierten Auswahl der Daten ist in Code 4.1 zu sehen. Zur Vereinfachung ist die Ausgabe von Informationen auf der Konsole und der Umgang mit Fehlern hier nicht dargestellt. Die als `source` bezeichneten Variablen repräsentieren die Pneumoniedaten. Variablen mit `target` im Namen stellen die Daten gesunder Patienten dar, auf welche die Eigenschaften der Pneumoniedaten projiziert werden. Der Methode nimmt durch den Parameter `strata` ein Array mit Spaltennamen entgegen, welche bei der Stratifikation berücksichtigt werden. Die Merkmale, die der Stratifikation übergeben werden, bezeichnet man als Stratum, beziehungsweise Strata. Die Datensätze werden anhand der übergebenen Strata gruppiert. Zwei verschachtelte `for`-Schleifen iterieren über diese Gruppen und konstruieren für jede Gruppe anhand der Strata eine Query. Diese Query wird auf die Daten der gesunden Patienten angewendet. Das Ergebnis der Datenabfrage wird anschließend anhand der Anzahl der Patienten mit Pneumonie in dieser Gruppe begrenzt.

Ein einfaches fiktives Beispiel verdeutlicht die Funktionsweise der Methode. Die Methode erhält die Daten gesunder und kranker Patienten und ein Array mit den folgenden Strata `['Projektion', 'Geschlecht']`. Angenommen es sind PA- und AP-Projektionen enthalten und das Geschlecht der Patienten teilt sich in weiblich und männlich, dann resultieren aus den Eigenschaften vier Gruppen.

- Projektion: PA, Geschlecht: weiblich
- Projektion: PA, Geschlecht: männlich
- Projektion: AP, Geschlecht: weiblich
- Projektion: AP, Geschlecht: männlich

Das Ergebnis der Query für die erste Gruppe enthält alle gesunden weiblichen Patienten, deren Bilder in PA-Projektion aufgenommen wurden. Wenn das Ergebnis beispielsweise 20 Einträge enthält, aber in den Pneumoniedaten nur 14 Bilder weiblicher Patienten in PA-Projektion existieren, werden die 20 Ergebnisse auf 14 Einträge begrenzt. Die Auswahl der Einträge geschieht dabei zufällig. Dieser Vorgang wird für die verbleibenden Gruppen wiederholt und deren Ergebnisse jeweils an die Ergebnisse der ersten Gruppe angehängt. Als Ergebnis gibt die Methode einen dataframe zurück, welcher gesunde Patienten enthält, deren Eigenschaften, denen der Patienten mit Pneumonie entsprechen.

Diese Vorgehensweise wurde auf jeden der vier großen Datensätze CXR14, CheXpert, MIMIC-CXR und PadChest angewendet. Open-i und Kermany beinhalten keine Metainformationen, die sich für die Stratifikation eignen. Eine Übersicht darüber, welche

```

1 def stratified_sampling_transferred(df_source, df_target, strata):
2     # Setting up local variables
3     source = df_source[strata]
4     target = df_target[strata]
5     source['size'] = 1
6     source_grpd = source.groupby(strata).count().reset_index()
7     target_grpd = target.groupby(strata).count().reset_index()
8
9     # Do both dataframes share the groups?
10    if not target_grpd[strata].equals(source_grpd[strata]):
11        ... # Logging + raising the error
12
13    # Calculate number of samples per stratification group
14    source_grpd['samp_size'] = round(source_grpd['size']).astype(int)
15
16    # Iterate over stratification groups
17    first = True
18    for i in range(len(target_grpd)):
19        # Iterate over each stratum to construct the query
20        qry=''
21        for s in range(len(strata)):
22            stratum = strata[s]
23            value = target_grpd.iloc[i][stratum]
24            if type(value) == str:
25                value = "!" + str(value) + "!"
26            if s != len(strata)-1:
27                qry = qry + stratum + ' == ' + str(value) + ' & '
28            else:
29                qry = qry + stratum + ' == ' + str(value)
30
31    # Apply query to dataframe
32    queried_df = df_target.query(qry)
33
34    # Is the number of retrieved samples sufficient?
35    n = source_grpd.iloc[i]['samp_size']
36    if len(queried_df) < n:
37        n = len(queried_df)
38        ... # Log warning
39
40    # Select n random samples of this group + store them
41    sampled_df = queried_df.sample(n=n).reset_index()
42    if first:
43        stratified_df = sampled_df
44        first = False
45    else:
46        stratified_df = stratified_df.append(sampled_df,
47                                              ignore_index=True)
48
49    return stratified_df

```

Code 4.1: Methode zur Stratifikation der gesunden Patienten, basierend auf den Eigenschaften der Patienten mit Pneumonie

	CXR14	CheXpert	MIMIC-CXR	PadChest
Projektion	✓	✓	✓	✓
Geschlecht	✓	✓	-	✓
Altersgruppe	✓	✓	-	✓
Medizinische Hilfsgeräte	-	✗	✓	-
Hersteller	-	-	-	✓
Röntgen-technologie	-	-	-	✓

- ✓ = Bei der Stratifikation berücksichtigt
- = Information nicht im Datensatz enthalten
- ✗ = Nicht bei der Stratifikation berücksichtigt

Tabelle 4.7: Übersicht der abhängig vom Datensatz berücksichtigten Eigenschaften bei der Stratifikation

Strata auf welchen Datensatz angewendet wurden, ist in Tabelle 4.7 zu sehen. Die Stratifikation von CheXpert musste ein Label als Stratum entfernen, da die Gruppen sonst so klein werden, dass die Datenmenge für die Anwendung der Stratifikation nicht mehr ausgereicht hätte. Projektion, Geschlecht und Altersgruppe wurden als wichtigste Eigenschaften identifiziert, weshalb das Label für medizinische Hilfsgeräte nicht berücksichtigt werden konnte.

Das Alter der Patienten ist zwar wertediskret, jedoch sorgt der große Wertebereich dafür, dass diese Eigenschaft im Gegensatz zu den anderen Eigenschaften in Tabelle 4.7 nicht als kategorische Eigenschaft gesehen werden kann. Somit eignet sich das Alter selbst nicht als Stratum. Da das Alter zumindest schätzungsweise anhand der Röntgenaufnahmen erkannt werden kann und die Inzidenz bei besonders jungen und besonders alten Personen sehr hoch ist, ist der Bias dieser Eigenschaft von besonderer Bedeutung. Daher wurden die je Datensatz drei Altersgruppen definiert, welche sich an charakteristischen Bereichen der Altersverteilung bei Patienten mit Pneumonie orientieren. Die Einführung der Altersgruppen ermöglicht es, das Alter in eine kategoriale Eigenschaft zu überführen, welche für die Verwendung als Stratum geeignet ist. Abbildung 4.11 zeigt, dass die Verwendung von Altersgruppen als Stratum die zuvor in Abbildung 4.10 dargestellten Unterschiede deutlich verringert. Die Auswirkung der Altersgruppen ist bei den restlichen Datensätzen sehr ähnlich, weshalb für diese keine weiteren Diagramme dargestellt sind. Die Verteilung der kategorischen Eigenschaften stimmen jeweils exakt überein, weshalb eine Darstellung dieser Verteilungen überflüssig ist.

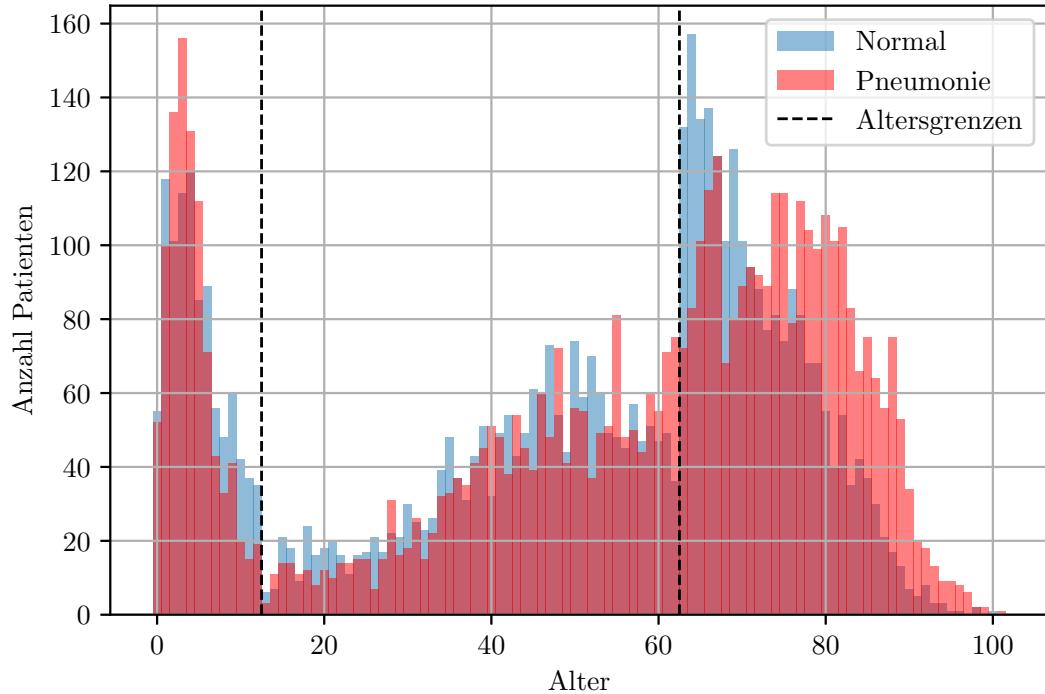


Abbildung 4.11: Altersverteilung der Patienten bei Einbeziehung der Altersgruppen im Stratifikationsprozess

4.3.3 Kombination der Datensätze

Die zuvor kennengelernte Methode der Stratifikation ist ein effektives Mittel, um den Bias in der Verteilung der Eigenschaften zu reduzieren. Da die Datensätze meistens nur Daten von einem Krankenhaus im Rahmen einer Kooperation erheben, ist die unzureichende Diversifikation der Daten unabhängig von der Stratifikation weiterhin präsent. Um die lokalen Gegebenheiten verschiedener Krankenhäuser in den Trainingsprozess einfließen zu lassen, bietet sich eine Kombination der stratifizierten Datensätze daher besonders an. Durch die Stratifizierung sind die spezifischen Eigenschaften innerhalb jedes Datensatzes möglichst ausgeglichen. Daher ist es möglich, die Datensätze beliebig zu kombinieren und unterschiedliche Tests durchzuführen.

Tabelle 4.7 hat gezeigt, dass die Eigenschaften Projektion, Geschlecht und Altersgruppe bei drei von vier Datensätzen als Strata angewendet wurden. Es wäre somit möglich gewesen, diese drei Datensätze zu fusionieren, um anschließend die Stratifikation durchzuführen. Bei dieser Vorgehensweise entstehen jedoch Abhängigkeiten zwischen den Datensätzen. Beim Entfernen oder Hinzufügen von Datensätzen würden die Proportionen der Eigenschaften somit aus dem Gleichgewicht gebracht werden. Die Verwendung einer der drei Datensätze als externer Validierungsdatensatz wäre somit nicht mehr möglich. Um die Flexibilität bei der Nutzung der Datensätze zu erhalten, wurden die Datensätze nicht fusioniert, sondern wie zuvor beschrieben in sich selbst stratifiziert. Die Kombination der Datensätze ist somit weiterhin möglich, aber nicht

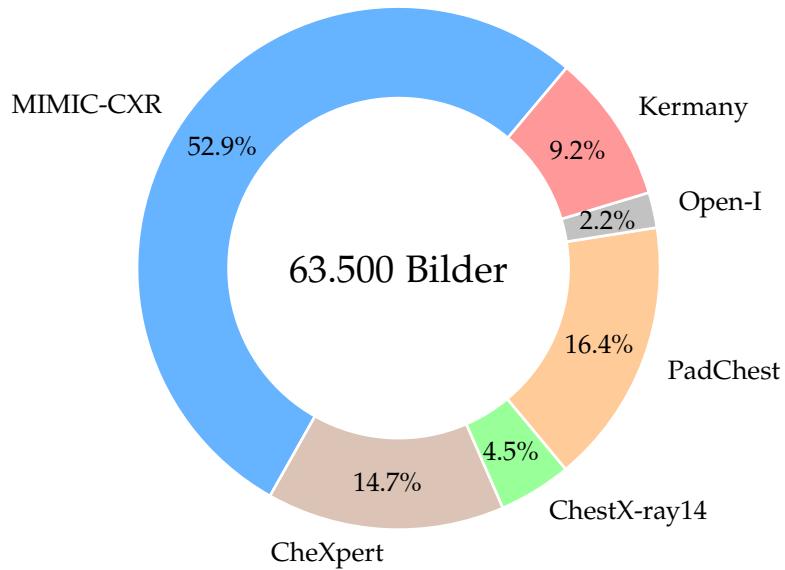


Abbildung 4.12: Prozentuale Anteile der Datensätze im Gesamtdatensatz

festgeschrieben.

Der Kermany Datensatz ist der Einzige der hier diskutierten Datensätze, der mehr Patienten mit Pneumonie, als Patienten ohne Pneumonie beinhaltet (siehe Abschnitt 4.2.3). Um keine wertvollen Pneumoniedaten zu verlieren, konnten gesunde Patienten von Open-i verwendet werden, um den Datensatz zu ergänzen. Die beiden Datensätze ergänzen sich somit gegenseitig, um eine Gleichgewichtung der Klassen sicherzustellen. Als ein weiterer positiver Aspekt ist eine Erhöhung der Varianz des Gesamtdatensatzes zu erwarten, da Open-i Röntgenbilder aus mehreren Quellen beinhaltet (siehe Abschnitt 4.2.1). Da der Kermany Datensatz keine Metainformationen zu den Bildern enthält, konnte bei der Kombination aus Kermany und Open-i keine Stratifikation durchgeführt werden.

Nach der in Abschnitt 4.2 durchgeführten Bereinigung und der im vorherigen Abschnitt erläuterten Stratifikation bleiben von initial etwa einer Millionen Röntgenbilder noch 63 500 übrig. Die Klassen sind dabei ausbalanciert, was einer Anzahl von je 31 750 Röntgenbildern pro Klasse entspricht. Im Vergleich dazu stehen für das Modelltraining nur wenige tausend Patienten mit Pneumonie zur Verfügung, wenn einer der hier beschriebenen Datensätze einzeln verwendet werden würde. Die prozentualen Anteile der Datensätze am Gesamtdatensatz sind im Kuchendiagramm in Abbildung 4.12 zu sehen. Da MIMIC-CXR besonders viele Pneumoniedaten enthält, ist dieser Datensatz im Vergleich zu den restlichen deutlich ausgeprägter.

Kapitel 5

Vorhersagemodell

Im vorherigen Kapitel wurden die Stärken und Schwächen der öffentlich verfügbaren Datensätze für Thoraxröntgenbilder analysiert und aus ihnen ein kombinierter Datensatz erstellt. Basierend auf diesen Daten soll nun ein Vorhersagemodell entwickelt werden, welches die Unterscheidung zwischen Patienten mit und ohne Pneumonie ermöglicht. Dabei werden aktuelle Erkenntnisse und Techniken angewandt, die sich in diesem Forschungsbereich bewährt haben.

Der Beginn dieses Kapitels erläutert theoretische Grundlagen, welche für das Verständnis der darauffolgenden Abschnitte erforderlich sind. Basiskenntnisse über den grundsätzlichen Aufbau und Trainingsprozess neuronaler Netze werden hierbei vorausgesetzt. Im darauffolgenden Abschnitt werden einige Schritte zur Datenvorverarbeitung entwickelt, welche die Qualität und Effizienz der Daten verbessern. Abschnitt 5.3 beschreibt die allgemeine Vorgehensweise, verwendete Techniken und Parametrisierung, die während des Modelltrainings angewendet wurden. In Abschnitt 5.4 wird die Vorgehensweise kritisch reflektiert, sowie Vor- und Nachteile des Modells in Bezug auf die Praxisrelevanz diskutiert. Zum Abschluss des Kapitels wird das entwickelte Vorhersagemodell mit dem aktuellen Stand der Technik verglichen und die erreichten Ergebnisse bewertet.

5.1 Theoretische Grundlagen

Dieser Abschnitt erläutert einige der in diesem Kapitel angewendeten Techniken, um ein besseres Verständnis der Modellentwicklung zu ermöglichen. Für grundlegende oder weiterführende Informationen über Deep Learning und den Trainingsprozess ist [90] besonders zu empfehlen.

5.1.1 Convolutional Neural Networks

Das CNN ist eine populäre und weit verbreitete Architektur tiefer neuronaler Netze. Ihre Hauptanwendung finden diese Netze vornehmlich in verschiedenen Disziplinen der Bildverarbeitung, wie beispielsweise der Objekterkennung oder Bildklassifikation.

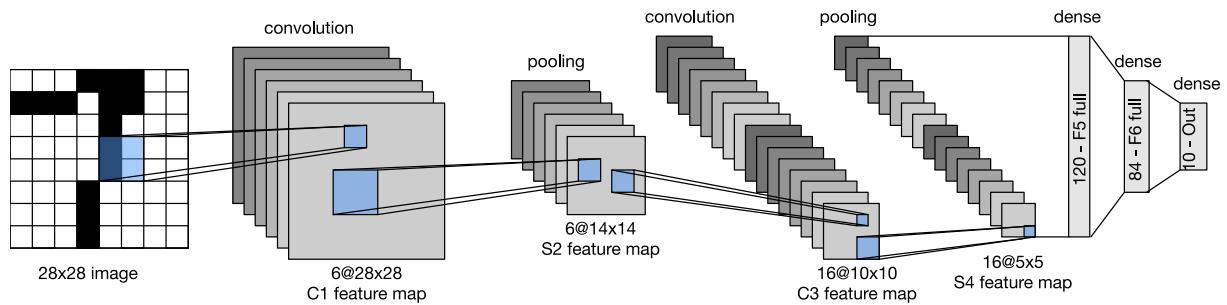


Abbildung 5.1: Visualisierung des Datenflusses von LeNet, einem der ersten erfolgreich eingesetzten CNNs. Die Eingabe ist eine handgeschriebene Ziffer, die Zielvariable ein Vektor mit Wahrscheinlichkeiten für zehn Ziffern. (Quelle: [90, S. 250])

Die Architektur wurde durch den Aufbau des visuellen Cortex im menschlichen Gehirn inspiriert, welcher für das Sehvermögen zuständig ist. Typischerweise bestehen CNNs aus mehreren Convolution- und Pooling-Schichten, gefolgt von einer oder mehreren Dense-Schichten. Abbildung 5.1 zeigt diesen grundsätzlichen Aufbau anhand eines Modells, welches zur Erkennung von handgeschriebenen Ziffern entwickelt wurde.

In der Convolution-Schicht wird eine diskrete Faltung durchgeführt, bei der ein Kernel schrittweise über die eingegebene Pixelmatrix geschoben wird. Das Modelltraining passt die Werte des Kernels dabei stetig an, um das Modell zu optimieren. Da die Höhe und Breite des Kernels im Normalfall größer als eins ist, verringert sich die Größe beider Achsen bei der Durchführung dieses Schritts [vgl. 90, S. 232].

Ähnlich wie bei der Faltung, schiebt die darauffolgende Pooling-Schicht ein Fenster über die Eingabematrix. Im Gegensatz zum Kernel beinhaltet dieses Fenster jedoch keine Parameter, sondern führen eine festgelegte Operation durch. Typischerweise berechnet die Pooling-Operation den Maximal- oder Durchschnittswert aller Elemente im Fenster. Ziel dieser Schicht ist die schrittweise Aggregation der zuvor von der Faltung berechneten Werte, um die Größe der Matrizen zu reduzieren [vgl. 90, S. 245f].

Die letzten Schichten des Modells sind in der Regel sogenannte Dense-Schichten, in denen alle Neuronen mit allen Ein- und Ausgabewerten verbunden sind. Bevor die mehrdimensionale Ausgabe der Convolution- und Pooling-Schichten an die Dense-Schicht weitergegeben werden kann, muss die Struktur zunächst in einen Vektor transformiert werden. Die letzte Dense-Schicht des Modells beinhaltet typischerweise für jedes vorherzusagendes Label ein Neuron, welches als Ausgabewert die Wahrscheinlichkeit der Zugehörigkeit zu diesem Label repräsentiert [vgl. 90, S. 250f].

5.1.2 Transfer Learning

In einigen Situationen kann es sinnvoll sein, eine bestehende Architektur mit vortrainierten Gewichten als Ausgangspunkt der eigenen Modellentwicklung zu verwenden. Der Überbegriff für diese Adaption eines Modells auf neue Daten und Aufgaben ist

Transfer Learning, da in diesem Prozess ein Wissenstransfer stattfindet. Eine Vielzahl der besten CNNs für Bildverarbeitungsaufgaben entstammen aus dem Wettbewerb ImageNet Large Scale Visual Recognition Challenge [74]. Der Wettbewerb stellt einen Datensatz mit 14 197 122 natürlichen Bildern zur Verfügung. Die neuronalen Netze treten beispielsweise in der Detektion und Lokalisation von Objekten gegeneinander an. Die besten Modelle des Wettbewerbs sind daher besonders als Ausgangspunkt für Transfer Learning geeignet. Das in Abschnitt 5.3.5 verwendete ResNet hat seinen Ursprung im ImageNet Wettbewerb [23].

Wenn eine Klassifikationsaufgabe nah genug an der Aufgabe eines vortrainierten Modells liegt, wird häufig das sogenannte Fine-Tuning, auf Deutsch etwa Feinabstimmung, eingesetzt. Unter der Annahme, dass das Modell in der Lage ist generelle Informationen aus Bildern zu extrahieren, werden alle Schichten des ursprünglichen Modells inklusive der gelernten Gewichte beibehalten. Einzig die letzte Dense-Schicht des Modells, welche die Zielvariablen der ursprünglichen Aufgabe enthält, wird durch eine Dense-Schicht mit neuen Zielvariablen ersetzt. Das anschließende Modelltraining optimiert lediglich die Gewichte dieser Schicht und führt somit die Namensgebende Feinabstimmung des Modells durch. Diese Vorgehensweise bietet sich besonders an, wenn nur eine kleine Menge Trainingsdaten für den Anwendungsfall vorhanden sind [vgl. 90, S. 578-583].

Weiterhin können vortrainierte Modelle zur Extraktion von Merkmalen verwendet werden, was auch als Feature-Extraction bezeichnet wird. Diese Technik schneidet das neuronale Netz beispielsweise nach einer Pooling-Schicht ab und ersetzt die abgeschnittenen Schichten durch eine eigens implementierte Architektur. Auf diese Weise können die auf den ursprünglichen Daten erlernten Konzepte als Eingabe für ein neues Modell verwendet werden. Ein Abschneiden nahe am Input des ursprünglichen Modells eignet sich tendenziell besser zur Extraktion von allgemeinen Merkmalen wie beispielsweise Kanten, während ein Abschneiden nahe dem ursprünglichen Output sich eher eignet, um globale Konzepte wie die Erkennung der erlernten Objektklassen zu übernehmen [vgl. 90, S. 643f].

5.1.3 Verwendete Bibliotheken

Die Datenvorverarbeitung und das Modelltraining verwenden einige externe Python Bibliotheken. Das Einlesen der Datensätze geschieht, ebenso wie deren Erstellung in Kapitel 4, unter Verwendung von pandas [81]. Die Vorverarbeitung der Röntgenbilder wurde mithilfe von Open Source Computer Vision Library (OpenCV) realisiert [57]. Die in Abschnitt 5.3.2 beschriebene Einteilung der Daten und die Auswertung der Klassifizierungsmaßnahmen verwenden die Bibliothek Scikit-learn [59]. Die Implementierung des Modells und des Trainingsprozesses nutzt die Bibliothek TensorFlow [1].

5.2 Datenvorverarbeitung

Kapitel 4 beschreibt die Eigenschaften der Datensätze und Metainformationen der Röntgenaufnahmen. Die Verfügbaren Informationen wurden genutzt, um eine Bereinigung der Daten durchzuführen, die Relevanz der Bilder sicherzustellen und die Verteilung der Eigenschaften möglichst gleichmäßig zu gestalten. Diese enorm Umfangreiche Analyse, Filterung und Kombination der Daten beinhaltet jedoch noch keine Vorverarbeitung der Bilddaten. Der nachfolgende Abschnitt erläutert die implementierten Vorverarbeitungsschritte, welche vor dem Modelltraining auf die Daten angewendet wurden, um deren Qualität zu verbessern.

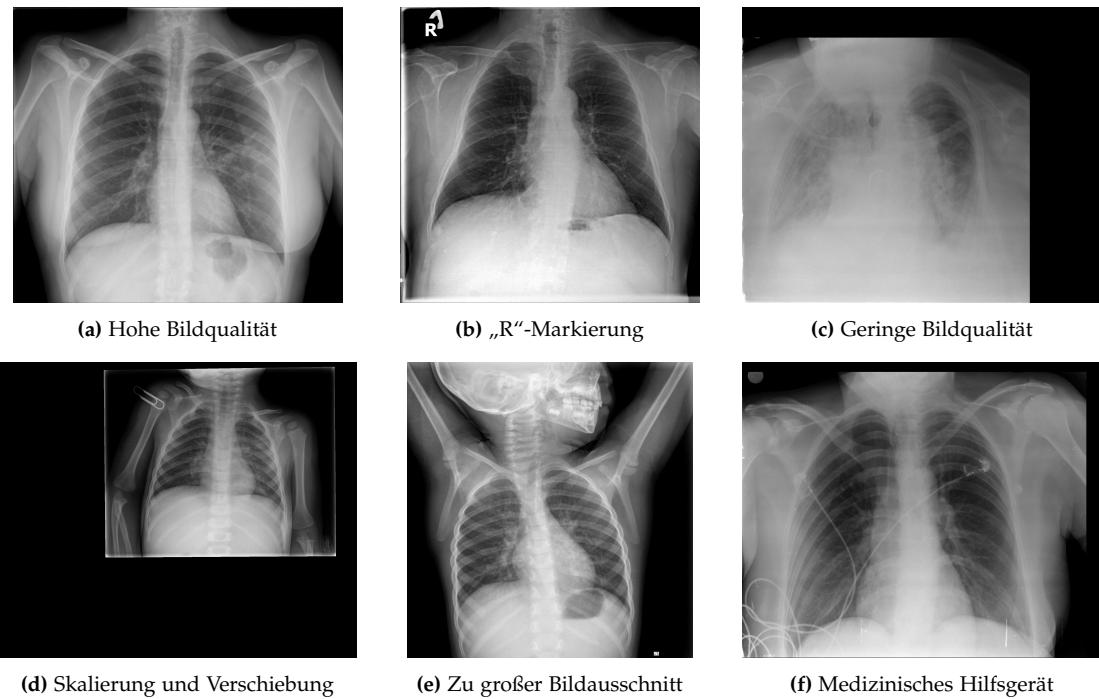


Abbildung 5.2: Die Diversität der Bilder sorgt für besondere Herausforderungen in der Datenvorverarbeitung (Quelle: [6])

Grundsätzlich unerwünschte Ansichten, wie beispielsweise laterale (siehe Abbildung 3.6b) oder einseitige Röntgenbilder (siehe Abbildung 4.7), sind während der Auswahl der Daten bereits erfolgreich entfernt worden. Somit beinhalten die Daten zu diesem Zeitpunkt ausschließlich frontale, vollständige Bilder. Abbildung 5.2 zeigt, wie unterschiedlich die Röntgenbilder trotz der ausführlichen Analyse der Metadaten noch sein können. Mangelhafte Qualität oder problematische Informationen innerhalb der Röntgenaufnahmen, wie sie in Abbildung 5.2b, 5.2c und 5.2f zu sehen sind, können in der Vorverarbeitung kaum optimiert werden. Durch die Nutzung der zuvor erwähnten Bibliothek OpenCV war es jedoch möglich, eine Verbesserung des Bildausschnitts und Seitenverhältnisses zu erreichen, was in den beiden nachfolgenden Abschnitten diskutiert wird. Zusätzlich implementiert Abschnitt 5.2.3 einen Vorverarbeitungsschritt zur Optimierung der Trainingsdauer.

5.2.1 Entfernen von Seitenrändern

In den Abbildungen 5.2c, 5.2d und 5.2f ist deutlich zu sehen, dass Randbereiche der Bilder schwarz sind und somit keine Informationen beinhalten. Weder die Präsenz, noch die Ursache dieser Verschiebung und Skalierung der Bilder wird in der Dokumentation der Datensätze diskutiert. Bei der manuellen Prüfung des kombinierten Datensatzes zeigte sich, dass zwar nur ein kleiner Teil der Bilder betroffen ist, es sich jedoch zweifelsfrei nicht um Einzelfälle handelt. Daher ist es möglich, eine Steigerung der Datenqualität durch die Korrektur dieser Bilder zu erreichen.

Der implementierte Vorverarbeitungsschritt entfernt diese Ränder, indem er die Konturen des eigentlichen Röntgenbildes detektiert und anschließend das kleinstmögliche Rechteck findet, das alle Bereiche der Kontur beinhaltet. Zur Erkennung der Konturen wird die Methode `cv2.findContours` verwendet. Die Methode ist für binäre Bilder implementiert, weshalb alle Pixel des Röntgenbilds zuvor mit dem Grauwert eins als Schwellwert binarisiert werden. Somit ist das Bild in schwarze und nicht-schwarze Pixel unterteilt. Die gesamten Konturen wird durch die Verwendung von `cv2.boundingRect` in Form eines Rechtecks ausgewählt. Der ausgewählte Bereich wird als vorverarbeitetes Bild dem nächsten Schritt übergeben und beinhaltet alle Pixel, welche sich innerhalb der rechteckigen Auswahlmaske befinden. Ein Vergleich zwischen Abbildung 5.3a und Abbildung 5.3b veranschaulicht die Auswirkung der Methode auf ein Röntgenbild.

5.2.2 Quadratisches Seitenverhältnis

Im Gegensatz zu den eher selten auftretenden schwarzen Seitenrändern, ist ein nicht-quadratisches Seitenverhältnis ein sehr präentes Problem in allen Datensätzen. Das in Abschnitt 5.3.5 beschriebene Modell erwartet als Eingabe ein quadratisches Bild. Sollte sich die Höhe von der Breite unterscheiden, muss vor dem Einspeisen in das neuronale Netz eine horizontale, beziehungsweise vertikale Stauchung oder Streckung des Bildes vorgenommen werden. Dieser Schritt sorgt für eine Verschiebung und Überlagerung der Bildinformationen.

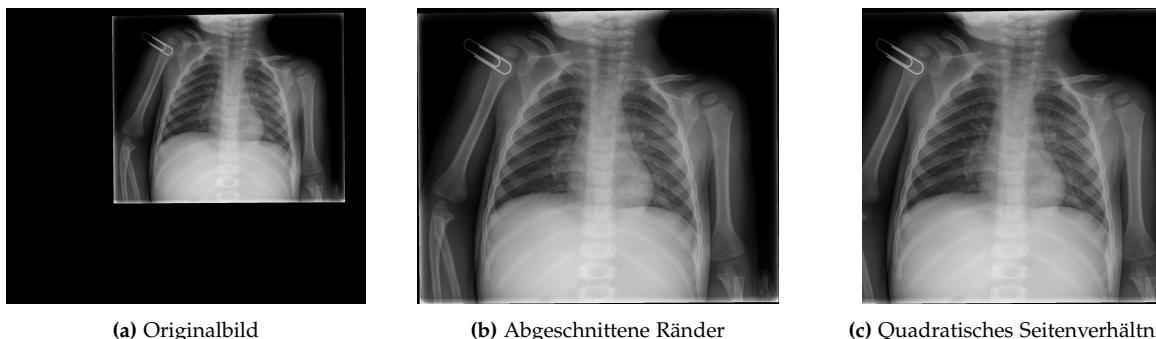


Abbildung 5.3: Korrektur der Ränder und Seitenverhältnisse der Röntgenaufnahmen als Vorverarbeitungsschritte

Besser als die Anpassung des Seitenverhältnisses durch Stauchung oder Streckung ist das zentrierte Abschneiden der horizontalen oder vertikalen Ränder. Die Empfehlung dieser Methodik ist in aktuellen Forschungsarbeiten dieses Themengebiets zu finden [vgl. 14, S. 3]. Die Verzerrung der Röntgenaufnahme ist schädlicher für das Vorhersagemodell als der potenzielle Informationsverlust durch das Zuschnieden.

Der nachfolgend beschriebene Vorverarbeitungsschritt behandelt ungleiche Seitenverhältnisse daher durch horizontales, beziehungsweise vertikales Zuschnieden der Bilder. Die Implementierung vergleicht die Höhe des Röntgenbildes mit der Breite, berechnet das Zentrum des Bildes und schneidet die längere der beiden Seiten passend zu. Für diese Methode war die Nutzung der Bibliothek OpenCV nicht notwendig. Der Vergleich von Abbildung 5.3b mit Abbildung 5.3c zeigt diesen Vorverarbeitungsschritt. Zudem wird in diesem beispielhaften Bild deutlich, dass die verworfenen Bereiche die Lungenregionen nicht betreffen und somit keine, für die Pneumonieklassifizierung relevanten, Informationen verloren gehen. Eine manuelle Überprüfung der Ergebnisse hat gezeigt, dass die Erhaltung relevanter Bildbereiche in fast allen Fällen gegeben ist. Der Teilverlust von Informationen über den Lungenbereich tritt nur unter sehr spezifischen Bedingungen auf, was die Empfehlung von Cohen u. a. bestätigt [14].

Röntgenaufnahmen ohne schwarzen Seitenrand, die sich bereits im quadratischen Seitenverhältnis befinden, bleiben während der ersten beiden Datenvorverarbeitungsschritte gänzlich unberührt. Ein Beispiel für ein solches Bild ist in Abbildung 5.2b zu sehen.

5.2.3 Skalierung der Bilder

Es wurde bereits angedeutet, dass das in Abschnitt 5.3.5 beschriebene Modell ein festgelegtes Format der Bilder voraussetzt. Dabei gibt die Implementierung des Modells nicht nur das Seitenverhältnis vor, sondern auch die Skalierung des Bildes. Als Eingabe werden Bilder in einer Auflösung von 224 mal 224 Pixel erwartet, was einer Gesamtanzahl von 50 176 Pixeln entspricht. Die Auflösungen der Röntgenaufnahmen variieren nicht nur zwischen den Datensätzen, sondern teils auch innerhalb des selben Datensatzes. Gemessen an der Anzahl der Pixel, übersteigen die Auflösungen der Röntgenbilder das erwartete Format mindestens um das 5-fache (kleinste Bilder aus Open-i [53]), bis hin zum 300-fachen (größte Bilder aus PadChest [6]).

Durch die Notwendigkeit der Skalierung jeder Röntgenaufnahme vor der Einspeisung in das CNN, entsteht ein enorm großer und wiederkehrender Rechenaufwand beim Modelltraining. Um diese redundanten Berechnungen zu vermeiden, wurde die Skalierung der Bilder gemeinsam mit den beiden zuvor beschriebenen Schritten vor dem Modelltraining durchgeführt, um die vorverarbeiteten Bilder anschließend separat abzuspeichern. Die Trainingsdauer konnte durch diese Vorgehensweise massiv reduziert werden. Abhängig von den verwendeten Daten und durchgeföhrten Tests konnte die Trainingsdauer teilweise um den Faktor 20 beschleunigt werden. Auch bei älteren Datensätzen wie CXR14 und Open-i, welche niedriger aufgelöste Bilder enthalten, konnte eine vielfache Reduktion der Trainingsdauer erzielt werden.

5.3 Modelltraining

Dieser Abschnitt beschreibt die beim Modelltraining eingesetzten Methodiken. Da die allgemeine Praxisrelevanz und Generalisierbarkeit im Vordergrund stehen, ist die umfangreiche Evaluation von Parameterkombinationen zur letzten Optimierung der Vorhersagegenauigkeit nicht Ziel dieser Arbeit. Viel bedeutender ist es, unter Verwendung bewährter Methoden, die Auswirkungen der kombinierten Datensätze mit reduziertem Bias zu analysieren, welche im vorherigen Kapitel detailliert beschrieben wurden. Bei der Auswahl der Vorgehensweisen orientiert sich diese Arbeit daher an anerkannten Forschungsarbeiten auf diesem Gebiet wie beispielsweise CheXNet, welches ein Vorhersagemodell für die Multi-Label-Klassifizierung von Thoraxröntgenbildern ist und von den Autoren des CheXpert Datensatz entwickelt wurde [66].

5.3.1 Verifikation des ursprünglichen Modells

Die in Abschnitt 1.1 erläuterte Vorgehensweise war der Ausgangspunkt für diese Arbeit. Ziel war es, diese Methodik auf den entwickelten Datensatz anzuwenden, um eine gesteigerte Generalisierung und Praxisrelevanz zu erreichen und die frühen Ergebnisse des Modells zu verifizieren. Das bereits trainierte U-Net Modell für die Lungensegmentierung (siehe Abbildung 1.1) wurde beibehalten, da die analysierten Datensätze keine Segmentierungen der Lungenbereiche beinhalteten und somit keine Möglichkeit bestand, neue Trainingsdaten einzuspeisen.

Um eine vergleichbare Ausgangssituation zu schaffen, wurden im ersten Schritt die Reproduktion der Ergebnisse vor dieser Arbeit angestrebt. Dafür wurde die gleiche Modellarchitektur erneut auf dem Kermany Datensatz trainiert. Unter Berücksichtigung kleiner Abweichungen war die Reproduktion der Ergebnisse erfolgreich. Bei der Betrachtung unterschiedlicher Durchgänge lag die Vorhersagegenauigkeit im Bereich von 90 % bis 95 %. Die Vorhersagegenauigkeit ermöglicht hier nur eine erste grobe Einschätzung des Modells. Ein aussagekräftigeres Qualitätsmaß ist die Area Under the Receiver Operating Characteristic Curve (AUROC), welche später in diesem Kapitel verwendet wird.

Aufgrund der geringen Datenmenge lag die Vermutung nahe, dass die hervorragenden Ergebnisse teilweise auf Überanpassung zurückzuführen sind. Diese These wurde überprüft, indem Röntgenbilder aus dem PadChest und dem CheXpert Datensatz in das Modell zur Vorhersage eingespeist wurden. Der Verdacht der mangelhaften Generalisierbarkeit bestätigte sich hierbei schnell, da die Vorhersagegenauigkeit lediglich im Bereich von 55 % bis 60 % lag. Ursächlich sind potenziell eine Überanpassung während des Trainings und die fundamentalen Unterschiede der Datensätze. Während der Kermany Datensatz beispielsweise nur Röntgenaufnahmen von Kindern bis fünf Jahren enthält, spiegeln PadChest und CheXpert sämtliche Altersbereiche wider.

Um dieser Problematik entgegenzuwirken, wurde das Modell anschließend auf dem in Abschnitt 4.3.3 beschriebenen kombinierten Datensatz trainiert. Der PadChest Da-

tensatz wurde dabei vom Training ausgeschlossen, um die Generalisierbarkeit auf externe Daten überprüfen zu können. Zusätzlich wurde ein weiterer Trainingsprozess mit den gleichen Daten durchgeführt, nur ohne die Lungensegmentierung zu verwenden. Durch diese Vorgehensweise kann die Wirksamkeit der zusätzlich zugeführten Lungenausschnitte überprüft werden.

Beim Vergleich der Ergebnisse zeigte sich, dass die Ergebnisse mit und ohne Lungensegmentierung fast identisch sind. Dieses Verhalten konnte auch für das Training auf beliebigen Kombinationen der Datensätze und auf einzelnen Datensätzen, außer dem Kermany Datensatz, repliziert werden. Somit führt die Lungensegmentierung leider im Allgemeinen nicht zu einer Verbesserung der Vorhersageergebnisse. Anhand dieser Erkenntnis lässt sich schlussfolgern, dass das Modell auch ohne die doppelte Gewichtung der Lungenbereiche in der Lage ist, relevante Regionen während des Trainings zu identifizieren. Aufgrund der ernüchternden Ergebnisse während der Verifikation ist die Beschreibung der Vorgehensweise und der Architektur hier und in Abschnitt 1.1 bewusst kurzgehalten.

5.3.2 Einteilung der Daten

Um die Vergleichbarkeit verschiedener Arbeiten und neuronaler Netze zu erleichtern, beinhalten die Datensätze häufig eine von den Autoren definierte Einteilung in Trainings-, Test- und Validierungsdaten. Beispielsweise enthält CheXpert separat abgelegte Validierungsdaten, welche keinen Patienten doppelt enthalten und manuell von Medizinern gelabelt wurden. Durch die höhere Qualität der Validierungsdaten und der verbesserten Vergleichbarkeit ist die Nutzung einer vorgegebenen Dateneinteilung somit sehr sinnvoll. Die in Abschnitt 4.3.3 beschriebenen Daten bestehen aus einer Kombination mehrerer Datensätze, weshalb die Nutzung vordefinierter Einteilungen an dieser Stelle nicht möglich ist. Daher war es notwendig eine eigene Unterteilung der Daten vorzunehmen. Für diesen Vorgang existiert keine im Allgemeinen ratsame oder standardisierte Vorgehensweise. Ein weit verbreiteter Ansatz, welcher hier zum Einsatz kommt, ist die Einteilung in 70 % Trainings-, 20 % Test- und 10 % Validierungsdaten. Wang u. a. verwendeten diese Verteilung bei der Evaluierung des CXR14 Datensatzes [vgl. 84, S. 7], ebenso wie Cohen u. a. bei der Entwicklung von Chester [vgl. 13, S. 3] und die Studie von Zech u. a. [vgl. 89, S. 5].

Die in Abschnitt 4.3.2 erläuterte Stratifizierung kann bei der Einteilung in Trainings-, Test- und Validierungsdaten ebenfalls hilfreich sein. Da bei der Auswahl der Daten bereits eine Gleichgewichtung der binären Zielvariable stattgefunden hat, ist eine Stratifizierung dieser Eigenschaft weniger bedeutsam. Wesentlich unausgeglichener und somit einflussreicher ist die Information über den Datensatz, aus dem das jeweilige Röntgenbild stammt. Die aus der Bibliothek Scikit-learn verwendete Methode `train_test_split` zur Einteilung der Daten bietet den Parameter `stratify` an, welchem eine Spalte des Dataframes übergeben werden kann, anhand dessen eine Stratifizierung durchgeführt wird. Damit eine Stratifizierung anhand der Zielvariable

und der Herkunft der Samples gleichzeitig möglich ist, wurde ein neues Feature, resultierend aus der Kombination beider Eigenschaften generiert. Die Anzahl der Klassen des generierten Features ergeben sich aus der Multiplikation der Anzahl möglicher Zustände der Zielvariable mit der Anzahl der Datensätze.

$$n_{\text{Kombiniert}} = n_{\text{Zielvariablen}} \cdot n_{\text{Datensätze}} = 12 \text{ Klassen} \quad (5.1)$$

Diese neu erzeugte Eigenschaft wird an den `stratify`-Parameter von Scikit-learn übergeben. Somit ist sichergestellt, dass die Balance zwischen Patienten mit und ohne Pneumonie, sowie die in Abbildung 4.12 gezeigten Anteile der Datensätze sich repräsentativ in den Trainings-, Test- und Validierungsdaten widerspiegeln.

5.3.3 Augmentation

Die Bildaugmentierung ist eine Technik aus dem maschinellen Sehen, welche zufällige Änderungen an Bildern vornimmt, um neue ähnliche Bilder zu generieren. Weit verbreitete Augmentationen sind beispielsweise die Rotation, Spiegelung, Translation und Skalierung. Die Technik wird häufig für die Veränderung der Trainingsdaten neuronaler Netze verwendet, um die Abhängigkeit zu bestimmten Bildeigenschaften zu reduzieren, wodurch die Generalisierbarkeit des Modells verbessert wird [90, S. 569-577].

Cohen u. a. trainierten mehrere CNN Modelle unter Verwendung des CXR14 Datensatz und nutzten sowohl beim Training als auch bei der Validierung, verschiedene Stufen von Augmentation [13]. Zusätzlich überprüften sie die Generalisierbarkeit der Modelle auf externen Daten, wofür sie den PadChest Datensatz nutzten. Die Ergebnisse zeigten, dass die Augmentation deutlich dazu beiträgt, eine bessere Generalisierbarkeit zu erreichen. Am besten schneidet das Modell bei zeitgleich angewendeten Augmentationen von bis zu 45° Rotation, 15 % Translation und 10 % Skalierung ab [vgl. 13, S. 14]. Die in Kapitel 4 beschriebenen Daten sind durch die Kombination der Datensätze bereits sehr divers. Daher verwendet diese Arbeit etwas geringer ausgeprägte Augmentationen, als gerade beschrieben. Als zufällige Augmentationen während des Trainings wurden bis zu 20° Rotation, 10 % Translation und 5 % Skalierung angewendet.

Die Autoren von CheXNet verwendeten für das Modelltraining zusätzlich eine zufällige horizontale Spiegelung [vgl. 66, S. 2]. Wie in Abschnitt 3.4.2 bereits erläutert wurde, werden Thoraxröntgenbilder per Definition so dargestellt, dass sich das Herz in der rechten Bildhälfte befindet. Eine horizontale Spiegelung als Augmentation würde diese Konvention verletzen und das Netz somit für ein irrelevantes Szenario trainieren. Diese Arbeit verzichtet daher auf eine zufällige horizontale Spiegelung.

5.3.4 Einfluss während des Trainings

Die zuvor beschriebene Machine Learning Bibliothek TensorFlow bietet die Möglichkeit, durch sogenannte Rückruffunktionen Einfluss auf den laufenden Trainingsprozess zu nehmen. Das in dieser Arbeit implementierte Modelltraining verwendet die

Rückruffunktionen `ReduceLROnPlateau`, `EarlyStopping` und `ModelCheckpoint`. Erste-
re ist so konfiguriert, dass sie die Verlustfunktion des Trainings überwacht und die
Lernrate des Optimierungsalgorithmus reduziert, wenn der Lernprozess über drei
Epochen stagniert. Wenn über einen Zeitraum von sechs Epochen keine Optimierung
der Ergebnisse mehr stattfindet, wird das Modelltraining durch die Rückruffunkti-
on `EarlyStopping` vorzeitig beendet. Somit hat die Reduktion der Lernrate genug
Epochen Zeit, um auf den Prozess einzuwirken, bevor das frühzeitige Stoppen des
Trainings auslöst. Weitere Informationen über die Handhabung der Lernrate während
des Modelltrainings können beispielsweise aus [90, S. 504-513] entnommen werden.
Die Rückruffunktion `ModelCheckpoint` überwacht zudem die Verlustfunktion in jeder
Epoche und speichert das beste Modell inklusive aller Gewichte.

5.3.5 Modellarchitektur

Der ursprüngliche Ansatz doppelt gewichteter Lungenregionen hat keine Verbesserung
der Vorhersagegenauigkeit erreicht. Daher war es notwendig, andere Modellarchi-
tekturen in Betracht zu ziehen. Ebenso wie in anderen Bereichen des maschinellen
Sehens liefert Transfer Learning hier die besten Ergebnisse und ist weit verbreitet
[14][37][66][62][89]. Die Stanford Machine Learning Group hat die Übertragbarkeit der
Ergebnisse der besten neuronalen Netze aus der ImageNet Challenge auf die Klassifi-
kation von Thoraxröntgenbildern untersucht [35]. Die Gruppe kam zu dem Ergebnis,
dass keine Korrelation zwischen der Vorhersagegenauigkeit in der ImageNet Chal-
lenge mit der Vorhersagegenauigkeit bei Röntgenbildern vorliegt. Bei diversen Tests
zeigte sich, dass es viel wichtiger für die Übertragbarkeit der Ergebnisse ist, welche
Architekturfamilie verwendet wird. Dabei schnitten die etwas älteren Architekturen
(ResNet, DenseNet) besser ab als die neueren (EfficientNet, MobileNet, MNASNet).
Dieses Ergebnis zeigt, dass etablierte Arbeiten wie [66] und [62] nach wie vor aktuell
sind.

ResNet-Architekturen erzielten in der zuvor erwähnten Untersuchung etwas bessere
Ergebnisse als die DenseNet-Familie [vgl. 35, S. 3ff]. Weiterhin basiert das führende
kommerzielle Assistenzsystem ebenfalls auf der ResNet-Architektur [vgl. 62, S. 4].
Daher wurde entschieden als Vorhersagemodell ein ResNet50 zu verwenden [23]. Die
letzte Dense-Schicht des Netzes wurde entfernt und durch eine Flatten-Schicht, gefolgt
von einer Dense-Schicht mit einem Neuron ersetzt, da in diesem Anwendungsfäll
nur eine Klasse vorherzusagen ist. Der Vergleich mehrerer Trainingsdurchgänge mit
unterschiedlich vielen eingefrorenen Schichten zeigte, dass das Modell die höchste
Genauigkeit erreicht, wenn alle Schichten und Gewichte zum Training freigegeben wer-
den, ohne dabei der Validierung auf externen Daten zu schaden. Das Modell verwendet
somit nicht Transfer Learning im engsten Sinne, sondern behält nur die Architektur
des Modells und die Initialisierung der Gewichte bei. Die vorliegenden Röntgenbilder
unterscheiden sich zudem sehr stark von den natürlichen Bildern in ImageNet, weshalb
typische Transfer-Learning-Techniken wie Fine-Tuning oder der Feature-Extraction
nicht angewendet werden können (siehe Abschnitt 5.1.2). Als Verlustfunktion verwen-
det das Modelltraining Binary Crossentropy und als Optimierungsalgoritmus das

stochastische Gradientenabstiegsverfahren. Sowohl die Verlustfunktion als auch der Optimierungsalgorithmus werden dabei von TensorFlow bereits zur Verfügung gestellt. Für die Normalisierung und Anpassung der Standardabweichung wurden die Werte der ImageNet Challenge verwendet. Für weitere nicht explizit erwähnte Parameter wie beispielsweise die Lernrate wurden die Standardwerte verwendet.

5.4 Bewertung des Modells

Dieser Abschnitt diskutiert die Eigenschaften und Ergebnisse des entwickelten Modells. Die ersten beiden Teilabschnitte beschäftigen sich mit der Bewertung der Vorhersageergebnisse im Vergleich zum aktuellen Stand der Technik und verifizieren die Praxistauglichkeit des Modells. Die verbleibenden Abschnitte setzen sich kritisch mit der generellen Vorgehensweise bei der Entwicklung von Vorhersagemodellen in diesem Forschungsbereich auseinander.

5.4.1 Vorhersagegenauigkeit

Die Vorhersagegenauigkeit wird anhand der Evaluierung interner und externer Daten ermittelt. Wie in Abschnitt 5.3.2 beschrieben, sind die Daten in 70 % Trainings-, 20 % Test- und 10 % Validierungsdaten unterteilt. Letztere werden nachfolgend auch als interne Validierungsdaten bezeichnet und wurden vor dem Modelltraining aus dem Datensatz entfernt. Damit auch die Vorhersagegenauigkeit bei komplett fremden Daten evaluiert werden kann, wurde der gesamte PadChest Datensatz vor dem Modelltraining entfernt. Wie in Abbildung 4.12 zu sehen ist trägt er 16,4 % zum Gesamtdatensatz bei und wird nachfolgend als externer Validierungsdatensatz verwendet.

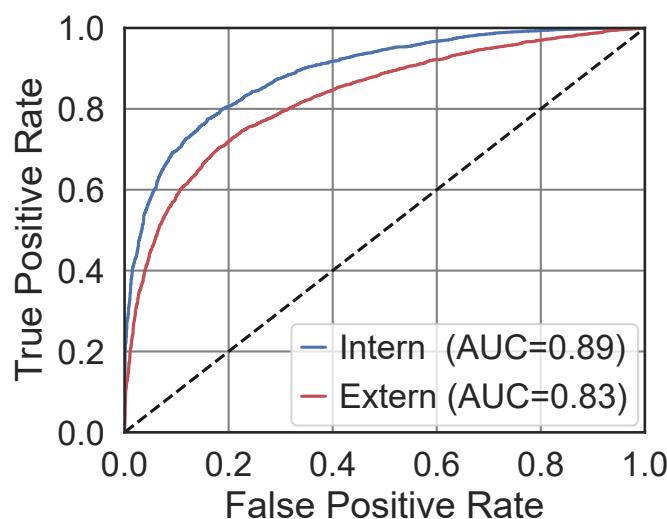


Abbildung 5.4: AUROC-Metriken des Vorhersagemodells, welches auf dem Gesamtdatensatz trainiert wurde. Sowohl bei den 10 % interner Validierungsdaten als auch bei PadChest als externe Daten zeigt das Modell eine gute Generalisierbarkeit.

Die Ergebnisse der internen und externen Evaluation können anhand der AUROC-Metrik in Abbildung 5.4 beurteilt werden. Mit dem Wert von 0,89 auf den internen Validierungsdaten spiegelt das entwickelte System den aktuellen Stand der Forschung wider, da aktuelle Arbeiten ähnliche Genauigkeiten angeben [89][62]. Die externen Validierungsdaten zeigen, dass auch Röntgenbilder aus unbekannten Quellen mit einer hohen Genauigkeit richtig diagnostiziert werden. Bei den in Abbildung 5.5 dargestellten Konfusionsmatrizen fällt jedoch auf, dass der Anteil falsch-negativer Vorhersagen bei PadChest deutlich höher ist als der Anteil falsch-positiver. Der Anteil korrekt identifizierter Krankheiten beziehungsweise Label wird auch als Sensitivität bezeichnet. In medizinischen Assistenzsystemen sollte die Sensitivität möglichst hoch sein, da falsch-negative Diagnosen gravierende gesundheitliche Folgen für den Patienten haben können.

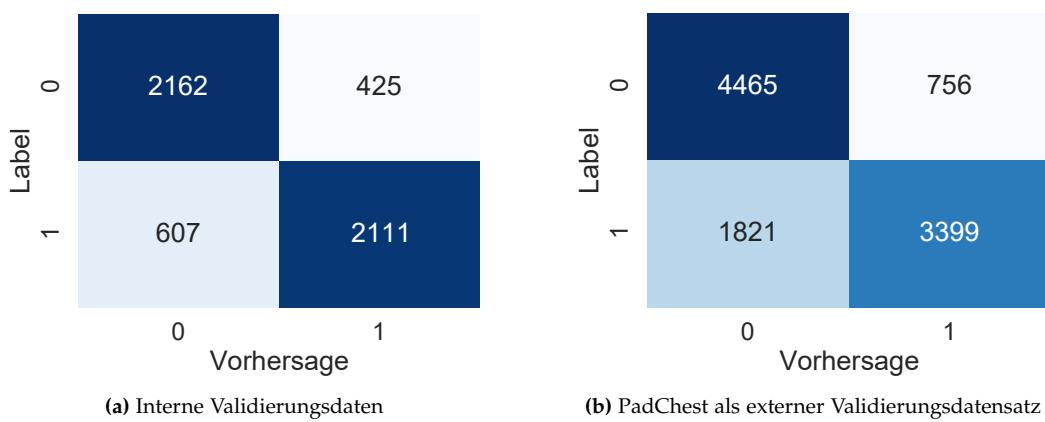


Abbildung 5.5: Konfusionsmatrizen für die Vorhersagen auf den internen und externen Validierungsdaten mit 1=Pneumonie und 0=Normal

5.4.2 Generalisierbarkeit

Wie im vorherigen Abschnitt bereits erwähnt wurde, ist die Genauigkeit auf den vom Training ausgeschlossenen Bildern des PadChest Datensatz nicht weit von der Genauigkeit der internen Validierungsdaten entfernt. Das System ist somit robust für die Analyse von Daten aus verschiedenen Quellen einsetzbar. Cohen u. a. befassten sich mit der institutionsübergreifenden Generalisierbarkeit bei der Klassifizierung von Röntgenbildern [14]. Sie kamen zu der Erkenntnis, dass es Grenzen für die Genauigkeit und Generalisierbarkeit der Vorhersagemodelle gibt, was auf darauf zurückzuführen ist, dass sich auch Mediziner bei der Befundung von Röntgenaufnahmen oft nicht einig sind. Dass bei der Erstellung der hier verwendeten Datensätze eine Mehrheitsabstimmung mehrerer Mediziner zur Vergabe der Labels notwendig war, verdeutlicht diese Situation.

Während der Datenanalyse in Abschnitt 4.2 wurde die ungleiche Verteilung der Eigenschaften besonders hervorgehoben und anschließend in Abschnitt 4.3 mehrere Konzepte zum Umgang mit dieser Situation entwickelt. Die bisher diskutierten Ergebnisse zeigen eine hervorragende Generalisierbarkeit des Vorhersagemodells. Ob das

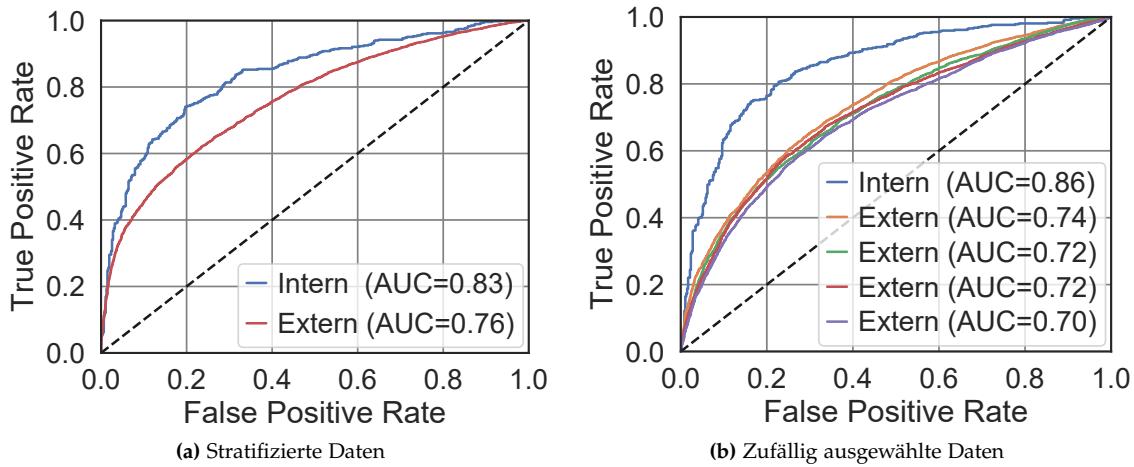


Abbildung 5.6: ROC-Kurven und AUROC-Werte mehrerer Vorhersagemodelle, die auf unterschiedlichen Teilmengen des CheXpert Datensatz trainiert wurden. Die Ergebnisse mit stratifizierten Daten zeigen eine bessere Generalisierbarkeit bei der Vorhersage externer Daten.

in Abschnitt 4.3.2 implementierte Verfahren zur Stratifikation der Eigenschaften zu diesen Ergebnissen beigetragen hat, wird nachfolgend verifiziert.

Als Erstes wurde ein Modell in der gleichen Architektur und Parametrisierung wie in Abschnitt 5.3.5 beschrieben trainiert. Als Trainingsdaten erhielt das Modell jedoch ausschließlich die stratifizierten Daten aus CheXpert. Ebenso wie in Abschnitt 5.4.1 wurden 10 % der Daten als interne Validierungsdaten und der stratifizierte PadChest Datensatz als externe Validierungsdaten verwendet. Die Ergebnisse dieses Trainings sind in der Receiver Operating Characteristic (ROC)-Kurve in Abbildung 5.6a zu sehen. Anschließend wurden anstelle der stratifizierten Daten mehrmals zufällige Daten aus dem vollständigen CheXpert Datensatz entnommen. Um auch hier ein ausgeglichenes Klassenverhältnis zu erhalten, wurde die Anzahl gesunder Patienten anhand der Anzahl der Pneumoniepatienten festgelegt. In diesen Daten befinden sich somit die Missverhältnisse der Eigenschaften des gesamten CheXpert Datensatzes, nur kleiner skaliert. Ein Vergleich der Abbildungen 4.11 und 4.10 demonstriert diese Unterschiede stratifizierter und nicht-stratifizierter Daten nach der Ausbalancierung der Klassen. Die Ergebnisse der zufällig ausgewählten Trainingsdaten sind in Abbildung 5.6b zu sehen. Da die Ergebnisse der internen Validierung in den Durchgängen nahezu identisch waren, wurde zugunsten der Übersichtlichkeit nur eine der ROC-Kurven für die interne Validierung dargestellt.

Der Vergleich der beiden Diagramme ermöglicht einige Erkenntnisse. So ist erkennbar, dass die Vorhersagen interner Validierungsdaten bei der Verwendung zufällig ausgewählter Daten besser abschneiden. Anhand der internen Validierung entsteht dadurch der Trugschluss, dass das Modell mit zufällig ausgewählten Daten eine höhere Genauigkeit und eine bessere Generalisierbarkeit bietet. Verifiziert man diese Ergebnisse jedoch anhand eines externen Datensatzes, wird schnell deutlich, dass der Grund für das bessere Abschneiden die Überanpassung an den Datensatz ist. Hier erreichen die stratifizierten Daten einen besseren AUROC-Wert als die zufällig ausgewählten Daten. Zusätzlich zeigen die stratifizierten Daten eine deutlich bessere Generalisierbarkeit der Ergebnisse, was anhand Differenz der AUROC-Werte, beziehungsweise des Abstands

der ROC-Kurven, zwischen internen und externen Validierungsdaten zu erkennen ist. Mehrere Durchgänge zeigen, dass die Überanpassung an den Datensatz abhängig von der willkürlichen Verteilung der Eigenschaften ist. Selbst bei einer sehr vorteilhaften Verteilung und somit geringeren Überanpassung sind zufällig ausgewählte Daten den stratifizierten Daten auf einem externen Datensatz unterlegen. Diese Evaluation zeigt, dass die in Abschnitt 4.3.2 entwickelte Stratifikation der Daten eine Reduktion des Bias und somit auch eine Steigerung der Generalisierbarkeit erreicht. Somit konnten wichtige Verbesserungen in Bezug auf die Fairness und Praxisrelevanz des Vorhersagemodells erreicht werden.

5.4.3 Zielvariable

Die Labeling-Prozesse der öffentlich verfügbaren Datensätze für Thoraxröntgenbilder ähneln sich sehr stark. In Kapitel 4 wurde ausführlich erläutert, wie die Erzeugung der Labels durch die Anwendung von NLP-Methoden auf Röntgenbefunde funktioniert. Diese Techniken zur Erzeugung der Labels und die zugrundeliegenden Informationen der Befunde sind bereits mehrfach in die Kritik geraten [55][56][75].

Eine Veröffentlichung von Pierson u. a. beschäftigt sich mit der Vorhersage des Schweregrads von Arthritis im Knie anhand von Röntgenbildern [61]. Die Label der verwendeten Daten basieren ebenfalls auf den Befunden und Einschätzungen der behandelnden Ärzte. Zusätzlich zu den Befunden der Ärzte wurden Daten über die Selbsteinschätzung des Schmerzempfindens der Patienten erhoben. Die Forscher entdeckten, dass die Einschätzungen der Ärzte im Vergleich zum Schmerzempfinden der Patienten einen rassistischen Bias enthalten. Das auf den Schweregrad der Arthritis trainierte CNN lernte diesen Bias und spiegelte ihn bei den Vorhersagen nicht-weißer Patienten wider. Die systematischen Fehler, welche sich in den medizinischen Diagnosen befinden, übertragen sich somit auf die trainierten Vorhersagemodelle. Diese Erkenntnis deckt sich mit den von Seyyed-Kalantari u. a. festgestellten Ungerechtigkeiten, die in den Datensätzen mit Thoraxröntgenbildern präsent sind [75]. Als Pierson u. a. das Modell anhand der vom Patienten angegebenen Schmerzwerte trainierten, konnte der rassistische Bias eliminiert werden. Daher ist es von zentraler Bedeutung die verwendeten Zielvariablen und Labels im medizinischen Kontext stets kritisch zu hinterfragen.

Bei der Anwendung dieser Erkenntnis auf das hier entwickelte Vorhersagemodell für Pneumonien stellt sich automatisch die Frage, ob es sinnvoll ist, Röntgenbefunde als zugrundeliegende Wahrheit für die Vergabe der Labels zu verwenden. Geeigneter als die Verdachtsdiagnose zum Zeitpunkt des Röntgenbefunds wäre die Information, welche Diagnose sich im weiteren Verlauf des Krankenhausaufenthalts bewahrheitet hat. Die Verdachtsdiagnose des initialen Röntgenbefunds könnte von dieser Information revidiert werden und sich somit sich als gänzlich falsch erweisen. Diese Art von Daten zu erheben ist enorm aufwändig, weshalb zum Zeitpunkt dieser Arbeit kein öffentlich verfügbarer Datensatz existiert, der diese Information beinhaltet. Zukünftige Arbeiten sollten die aus den Labels resultierenden Zielvariablen daher weiter

kritisch hinterfragen und alternative Konzepte zur Einteilung der Krankheitsklassen untersuchen.

5.4.4 Multi-Label-Klassifizierung

Die in Abschnitt 4.2 beschriebenen Datensätze enthalten, abgesehen von Kermany und Open-i, viele unterschiedliche Krankheitslabel. Häufig werden daher neuronale Netze zur Multi-Label-Klassifizierung implementiert, welche als Vorhersage keinen einzelnen Wert, sondern einen Vektor mit Vorhersagewerten ausgeben.

Aufgrund der leicht voneinander abweichenden Labeling-Prozesse überschneiden sich nicht alle Labels der unterschiedlichen Datenquellen. Des Weiteren können zwischen den Datensätzen Abweichungen der Definition eines Labels auftreten, sodass selbst bei der Überschneidung des Labels der Informationsgehalt nicht identisch ist. Zusätzlich zur besseren Vergleichbarkeit mit anderen Arbeiten ist dies ein Grund, weshalb die meisten Forschungsarbeiten nur einen Datensatz verwenden. Da alle Datensätze ein eindeutig definiertes Krankheitslabel für Pneumonien besitzen und sich diese Arbeit auf die Klassifizierung von Pneumonien beschränkt, ist diese Arbeit von dieser Problematik nicht betroffen.

Für den praktischen Einsatz des Modells ist eine Multi-Label-Klassifizierung sehr interessant. Die behandelnden Mediziner stellen während des Diagnoseprozesses verschiedene Verdachtsdiagnosen gegenüber und wägen ab, welche der Diagnosen die wahrscheinlichste ist (siehe Abschnitt 3.1). Die Erweiterung des hier entwickelten Modells für die Multi-Label-Klassifizierung wäre daher ein denkbarer Schritt für zukünftige Arbeiten. In diesem Fall ist zu beachten, dass die Implementierung des in Abschnitt 4.3.2 erläuterten Stratifikationsverfahrens angepasst werden muss. Die Umsetzung des Algorithmus wird dabei komplexer, da die Dimension der zu berücksichtigenden Eigenschaften steigt. Da Lungenentzündungen enorm häufig auftreten, kann die Vorhersage des entwickelten Modells dennoch bereits von großer Bedeutung sein und als schnell verfügbare zweite Meinung einen Mehrwert im Diagnoseprozess bieten.

Kapitel 6

Grafische Benutzeroberfläche

Der Anwendungsbereich des entwickelten Vorhersagemodells befindet sich im medizinischen Sektor. Folglich sind die Nutzer des Systems hauptsächlich Ärzte und Radiologen, die in der Regel nur über begrenzte Informatikkenntnisse verfügen. Daher wurde für die Nutzung des Vorhersagemodells eine intuitive Benutzeroberfläche entwickelt, die in diesem Kapitel vorgestellt wird. Zu Beginn wird der Funktionsumfang und die Zielsetzung des Systems erklärt. Nachfolgend werden die Technologien vorgestellt, die bei der Entwicklung eingesetzt wurden. Anschließend wird der visuelle Aufbau und die Nutzung des CAD-Systems demonstriert. In den letzten beiden Abschnitten werden die Hauptkomponenten der Anwendung erläutert, sowie die Stärken und Schwächen des entwickelten Systems durch den Vergleich mit anderen Lösungsansätzen diskutiert.

6.1 Ziel und Funktionsumfang

In diesem Abschnitt wird die Motivation und das Ziel der Benutzeroberfläche definiert, sowie die dazugehörigen Anforderungen formuliert. Die Anforderungen sind dabei nicht als Vorgaben für einen produktiven Einsatz gedacht, sondern als Unterstützung zur Entwicklung eines möglichst guten Prototyps. Eine prototypische Umsetzung ist für die Zielsetzung dieser Arbeit ausreichend, da das System für die Demonstration der Technologie und den fachlichen Austausch mit Medizinern gedacht ist. Die Anwendung soll unkomplizierte Experimente ermöglichen und ein Verständnis für die Fähigkeiten neuronaler Netze zu schaffen. Daher wurde auf die umfangreiche Definition von Software-Anforderungen, wie beispielsweise in [85] beschrieben, bewusst verzichtet.

Die Nutzungsart der Benutzeroberfläche kann sehr vielfältig sein, daher sollte die Anwendung weitgehend unabhängig vom Betriebssystem oder der verwendeten Hardware sein. Um einen unabhängigen Betrieb der Anwendung vom jeweiligen Endgerät zu ermöglichen, wurde beschlossen die Benutzeroberfläche in Form einer Webanwendung zu realisieren. Da die Anwendung hauptsächlich im akademischen Bereich zum Einsatz kommt, wurde sie in englischer Sprache entwickelt.

6.1.1 Zielgruppe

Bei der Entwicklung einer Anwendung sollte beachtet werden, für welche Nutzer die Anwendung gedacht ist. Die hier beschriebene Benutzerschnittstelle verfolgt zwei Hauptziele und hat somit auch zwei Zielgruppen.

1. Visualisierung der Vorhersagen für Demonstrationszwecke
2. Prototypische Umsetzung eines CAD-Systems

Das erste Szenario, die Visualisierung der Vorhersage, beinhaltet als Zielgruppe prinzipiell jeden, der an der Anwendung von KI im medizinischen Bereich interessiert ist. Abhängig vom Personenkreis einer Demonstration kann die Zielgruppe hierbei natürlich variieren. Die Darstellung der Ergebnisse sollte daher leicht verständlich und intuitiv sein, damit die Vorhersage des Modells auch ohne medizinische Vorkenntnisse oder spezielles Wissen über neuronale Netze interpretierbar ist. Bei einer Demonstration können dadurch auch Laien die verschiedenen Verhaltensweisen des Systems verstehen. Besonders gute oder problematische Beispiele von Vorhersagen sollen zudem leicht gegenübergestellt und diskutiert werden können.

Betrachtet man die grafische Benutzeroberfläche in ihrer Funktion als CAD-System, resultiert daraus eine rein medizinische Zielgruppe, bestehend aus Ärzten, Radiologen und anderem medizinischem Fachpersonal. Durch die mediale Aufmerksamkeit können KI-Systeme für Außenstehende mit Vorurteilen belastet sein. Ähnlich wie bei Chester [13] ist die Intention der Anwendung in diesem Kontext daher, Aufklärungsarbeit zu leisten und eine Schnittstelle zwischen Medizinern und Informatikern zu schaffen. Umgekehrt kann die Resonanz aus dem medizinischen Bereich Aufschluss darüber geben, wie ein echter Mehrwert im Diagnoseprozess geschaffen werden kann und welche Funktionalitäten eher als technische Spielereien betrachtet werden können. Der Fokus liegt somit auf der Entwicklung einer prototypischen CAD-Systems, welches nicht für den produktiven Einsatz gedacht ist, sondern durch die praxisnahe Visualisierung des Vorhersagemodells die Kommunikation zwischen den Domänen erleichtern soll.

6.1.2 Selbsterklärendes System

Wie eingangs erwähnt wurde, soll die Bedienoberfläche möglichst intuitiv sein, um eine breite Zielgruppe anzusprechen. Die Verwendung von Standardkomponenten für Webseiten unterstützt dieses Ziel, da diese bereits bekannt sind und sich die Bedienung vertraut anfühlt. Um Missverständnissen vorzubeugen und über die Intention und Nutzung der Anwendung zu informieren, bietet sich die Entwicklung einer Startseite besonders an. Der Nutzer muss zudem darüber aufgeklärt werden, dass es sich bei der Anwendung um einen experimentellen Prototyp handelt, der weder medizinisch noch kommerziell eingesetzt werden darf.

Da Intuition subjektiv sein kann, ist ein Tutorial Bereich zu erarbeiten, der die Nutzung der Anwendung erklärt. Zusätzlich kann dieser Bereich dafür genutzt werden, weitere

Informationen über die Verarbeitung der hochgeladenen Daten zur Verfügung zu stellen. Für Nutzer, denen keine Thoraxröntgenbilder zur Verfügung stehen, soll zudem eine Möglichkeit gefunden werden, Beispielbilder in das System einzuspeisen.

6.1.3 Analyse von Röntgenbildern

Die wohl wichtigste Funktionalität der Benutzeroberfläche ist die Kommunikation mit dem Vorhersagemodell. Es muss daher eine Schnittstelle geschaffen werden, um Röntgenbilder in das System einzuspeisen und die Ergebnisse anschließend darzustellen. Es wird vorausgesetzt, dass die zu analysierenden Röntgenbilder als lokale Dateien verfügbar sind und sich in einem geeigneten Dateiformat befinden. Das System sollte dabei kompatibel mit den Formaten JPEG [83] und Portable Network Graphics (PNG) [29] sein.

Nach der Auswahl des Röntgenbilds soll dieses in der grafischen Oberfläche dargestellt werden, damit das Ergebnis des Algorithmus visuell überprüft werden kann. Die vorhergesagten Klassen könnten beispielsweise anhand einer Skala dargestellt werden, die eine grobe Aussage darüber trifft, wie sicher sich das Modell bei der Entscheidung war. Somit können Entscheidungen im Grenzbereich kritischer hinterfragt werden. Weiterhin soll es möglich sein, mehrere Analysen darzustellen, um die verschiedenen Szenarien miteinander vergleichen zu können. Ein paralleles Hochladen und Analysieren mehrerer Bilder ist dabei nicht notwendig, da die Beurteilung eines Bildes durch das Modell nur wenige Sekunden in Anspruch nimmt.

6.1.4 Optimierung für mobile Endgeräte

Abhängig vom Verwendungszweck kann es sinnvoll sein, die Anwendung von einem mobilen Endgerät wie einem Smartphone oder Tablet aus aufzurufen. Denkbare Szenarien hierfür sind beispielsweise das Abfotografieren und Hochladen eines zu analysierenden Röntgenbilds oder die Präsentation der Untersuchungsergebnisse beim Patienten. Um die Flexibilität der Anwendung zu demonstrieren, sollte diese Funktionalität bereits im Prototypen enthalten sein. Eine automatische Anpassung des Layouts bei der Verwendung von mobilen Geräten ist daher erforderlich. Die visuellen Komponenten sollten dem Nutzer ebenso wie in der Desktopansicht vertraut erscheinen. Die zuvor in Abschnitt 6.1.2 definierte Anforderung zur Verwendung von Standardkomponenten erweitert sich somit auf die Verwendung von Komponenten, die auch in mobilen Anwendungen verbreitet sind.

6.1.5 Demonstration der Erweiterbarkeit

Es wurde bereits erwähnt, dass die Anwendung demonstrieren soll, welche Möglichkeiten die Technologie, beziehungsweise der gewählte Ansatz bieten. Da eine

Webanwendung mühelos im laufenden Betrieb erweitert werden kann, sollte beispielhaft eine Funktionalität integriert werden, die nicht direkt mit dem Analysieren der Röntgenbilder zusammenhängt. Somit kann demonstriert werden, dass der Ansatz sehr viel Flexibilität und Anpassungsmöglichkeiten bietet. Weiterhin soll die Webanwendung über eine zusätzliche, simulierte Funktionalität verfügen, um zu demonstrieren, dass die Applikation für weitere Aufgaben eingesetzt werden könnte.

6.2 Verwendete Technologien

Dieser Abschnitt erläutert kurz die Technologien, die bei der Entwicklung der Webanwendung verwendet wurden. Als Webanwendungs-Framework wurde Angular gewählt, die Bedienelemente sind aus dem Material Design System von Google und für spezielle Visualisierungen bei der Analyse von Röntgenbildern wurden zwei externe Bibliotheken verwendet.

6.2.1 Angular Framework

Zu Beginn von Abschnitt 6.1 wurde bereits festgelegt, dass die Anwendung als Webanwendung umgesetzt werden sollte. Aufgrund ihres Funktionsumfangs und architektonischer Vorteile sind Single-Page-Webanwendungen heutzutage sehr verbreitet. Unter einer Single-Page-Webanwendung versteht man eine Webseite oder -anwendung, die aus einem einzigen HTML-Dokument besteht und weitere Elemente oder Ansichten dynamisch nachlädt. Durch die Nutzung dieser Architektur sind die Webanwendungen eigenständiger, können besonders dynamische Inhalte anbieten und verringern die benötigte Client-Server-Kommunikation.

Daher wurde die vorliegende Webanwendung als Single-Page-Webanwendung konzipiert, wobei als Entwicklungsplattform Angular verwendet wurde [20]. In Angular wird das Layout in HTML, das Styling in Cascading Style Sheets (CSS) und die Programmlogik in TypeScript entwickelt. Benötigte Funktionalitäten können durch TypeScript-Bibliotheken eingebunden werden, wobei Kernfunktionalitäten durch interne Bibliotheken des Frameworks bereits zur Verfügung gestellt werden. Angular ist für mobile Geräte und Desktop gleichermaßen geeignet und bietet Möglichkeiten, die Inhalte flexibel an das jeweilige Endgerät anzupassen. Somit konnte die in Abschnitt 6.1.4 definierte Anforderung zur Optimierung für mobile Endgeräte erfüllt werden.

Weiterhin wurde Material Design verwendet, welches ein anpassungsfähiges System zur Erstellung visueller Komponenten darstellt [22]. Ebenso wie Angular wird Material Design unter der Führung von Google entwickelt, weshalb eine spezielle Integration des Designsystems für Angular existiert [21]. Besonders häufig wird Material Design bei der Entwicklung von Android Anwendungen und Webanwendungen verwendet. Die Verwendung des Material Stils und der Standardkomponenten sollte demnach für viele Anwender eine vertraute Umgebung schaffen.

6.2.2 Externe Bibliotheken

Als externe Angular Bibliotheken wurden *Angular Gauge Chart* [68] und *NgxSpinner* [11] verwendet. Das Gauge Chart wird eingesetzt, um die Klassifikationsergebnisse des Vorhersagemodells zu visualisieren. Die Ausgabe des Vorhersagewertes auf einer Skala ermöglicht dem Nutzer zu erkennen, wie deutlich das Modell zu einer Klasse tendiert. *NgxSpinner* wird verwendet, um bei Serveranfragen eine Animation anzuzeigen, bis die Antwort des Servers eingetroffen ist.

6.3 Aufbau der Anwendung

Dieser Abschnitt beschreibt die Architektur und den visuellen Aufbau der Webanwendung. Die implementierten Komponenten und ihr Zusammenspiel werden vorgestellt, sowie die Modularität und Erweiterbarkeit des Systems erläutert. Für die Erklärung des visuellen Aufbaus wird jede Seite, auch View genannt, in einem separaten Abschnitt beschrieben. Durch die Verwendung von Bildschirmaufnahmen wird weiterhin erklärt, welche Funktionalitäten dem Nutzer zur Verfügung stehen und wie Bedienung dieser abläuft.

6.3.1 Architektur

Da die Webanwendung auf Angular basiert, ist sie als Single-Page-Anwendung konzipiert. Somit gibt es ein Hauptdokument, welches weitere Inhalte dynamisch nachlädt. Weiterhin bietet das Angular Framework einige Konzepte, die eine modulare Gestaltung der Anwendung erleichtern und somit die Wiederverwendbarkeit implementierter Komponenten fördern. Die hier verwendeten Konzepte sind Views, Components, Services und Routing, welche nachfolgend im Detail erläutert werden. Wenn in den folgenden Absätzen Begriffe *kursiv* dargestellt sind, handelt es sich um Komponenten des in Abbildung 6.1 dargestellten Architektschaubilds.

Das Hauptdokument ist die Basis der Anwendung und wird beim Öffnen der Anwendung als erstes aufgerufen. Von hier aus werden Views und Components je nach Bedarf dynamisch geladen. Um durch die Applikation navigieren zu können, wird eine Navigationsleiste auf der linken Seite der Anwendung verwendet, welche für jede View eine Verknüpfung bietet. Die Component, in der diese Leiste umgesetzt wurde heißt *NavigationBar* und wird direkt vom Hauptdokument genutzt. Somit ist die Navigationsleiste unabhängig von der aktuell dargestellten View und wird immer angezeigt.

Um zwischen den Views wechseln zu können, wird das Angular-Konzept des Routers angewendet. Das Routing-Modul befindet sich im Hauptdokument und tauscht den Inhalt der Seite aus, wenn beispielsweise durch die Navigationsleiste auf eine andere View zugegriffen werden soll. Somit ist es möglich, nur die Daten nachzuladen, die

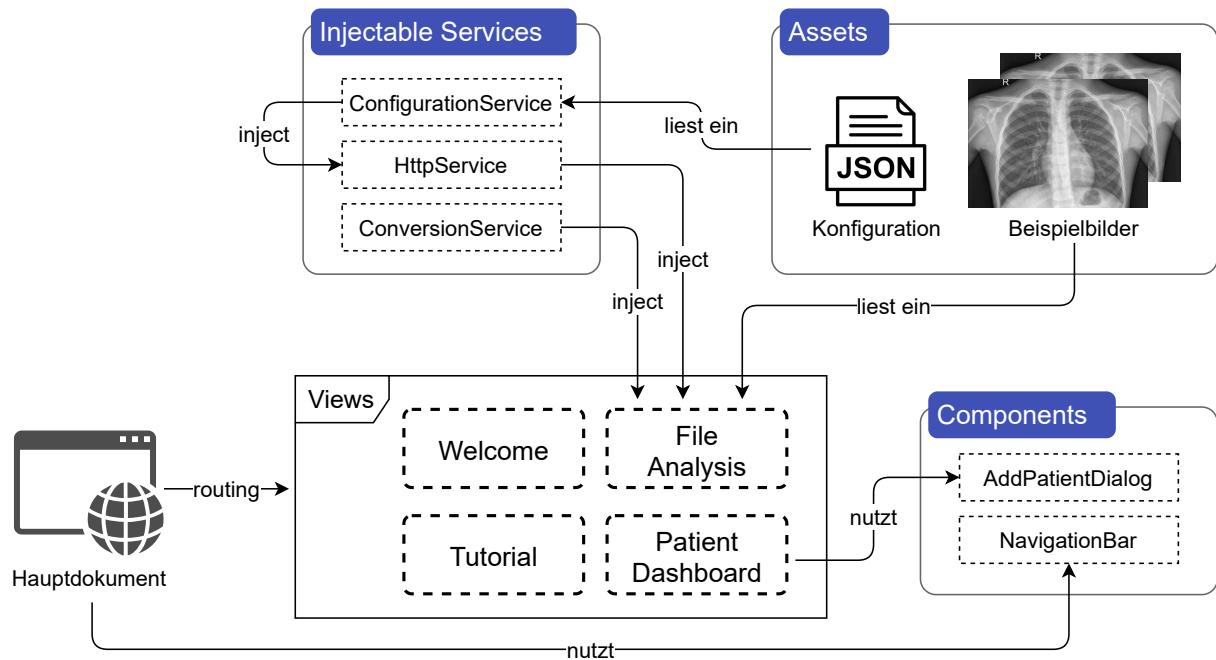


Abbildung 6.1: Übersicht und Zusammenspiel der Komponenten der Webanwendung

aktuell benötigt werden, ohne das gesamte Hauptdokument neu laden zu müssen. Der Router ist so konfiguriert, dass beim Aufruf der Webanwendung automatisch die *Welcome*-View geladen wird, auch wenn sie nicht Teil des aufgerufenen Hauptdokuments ist.

Um Funktionalitäten zu kapseln und wiederverwendbar zu machen, werden Services eingesetzt. Services implementieren ausschließlich Programmlogik, die nicht View-spezifisch ist und bei Bedarf überall zur Verfügung stehen kann. Zusätzlich zur Modularität und Wiederverwendbarkeit hilft dieser Ansatz dabei, die Logik der Views effizient und übersichtlich zu halten. Wenn ein Service eingebunden und genutzt wird, spricht man auch von Dependency Injection (DI). Dabei können die Services sowohl in Views und Components als auch in andere Services injiziert werden.

Im Gegensatz zu Services beinhalten Components zusätzlich zur TypeScript Programmlogik auch immer einen visuellen HTML Anteil. Components sind ebenfalls wiederverwendbar und können durch andere Components genutzt werden. Typische Beispiele für flexibel einsetzbare Components sind Eingabeaufforderungen oder Benachrichtigungen, die prinzipiell in jeder View darstellbar sein müssen. Im Angular Framework werden die Seiten der Anwendung ebenfalls als Components entwickelt. Umgekehrt betrachtet ist jedoch nicht jede Component auch eine vollwertige View.

Die Views *Welcome* und *Tutorial* beinhalten statische Informationen, weshalb sie keine weiteren Components oder Services nutzen. Die Seite *Patient Dashboard*, welche nachfolgend in Abschnitt 6.3.4 beschrieben wird, bietet durch das Einbinden der Component *AddPatientDialog* die Möglichkeit, einen neuen Patienten zur Übersicht hinzuzufügen. Diese Funktionalität wurde in Form einer Component realisiert, da diese nicht exklusiv

```

1 | {
2 |     "backend": {
3 |         "stage1_url": "http://127.0.0.1:5000/",
4 |         "stage2_url": "http://127.0.0.1:5001/",
5 |         "stage3_url": "http://127.0.0.1:5002/"
6 |     }
7 |

```

Code 6.1: Beispielhafte Konfigurationsdatei der Webanwendung im JSON-Format, wobei sich die Zieladresse des Servers auf dem lokalen System befindet

der Patientenübersicht zur Verfügung stehen sollte.

Den wohl wichtigsten Teil der Anwendung stellt die View *File Analysis* dar. Hier können die Röntgenbilder hochgeladen und analysiert werden. Um eine Verbindung mit dem Backend aufzubauen zu können, wurden die Services *ConfigurationService* und *HttpService* implementiert. Zusätzlich zur Klassifizierung von Pneumonien, wird in der View *File Analysis* eine Unterscheidung zwischen bakteriellen und viralen Pneumonien angedeutet. Die Information über bakterielle und virale Pneumonien ist lediglich im Kermany Datensatz enthalten, weshalb die Datenmenge für diesen Anwendungsfall verschwindend gering ist (siehe Abschnitt 4.2.3). Um mehr als nur eine reine Simulation der Funktionalität anbieten zu können, wurde ein zweites Vorhersagemodell für diese Unterscheidung anhand des Kermany Datensatzes trainiert. Das Modell hat jedoch keinerlei praxisrelevante Aussagekraft und ist als reine Simulation der Funktionalität für Demonstrationszwecke anzusehen, weshalb es in Kapitel 5 bewusst nicht beschrieben wurde. Die Demonstration der Erweiterbarkeit der Webanwendung für weitere Aufgaben, wie sie in Abschnitt 6.1.5 gefordert wurde, konnte somit erfüllt werden.

Der *ConfigurationService* liest eine Konfigurationsdatei ein, welche Informationen über die Adressen und Ports des Backend enthält (siehe Code 6.1). Die Datei befindet sich dabei im JavaScript Object Notation (JSON)-Format, welches einfach lesbar ist und sich für das Speichern strukturierter Daten besonders gut eignet. Da das Laden lokaler Dateien in TypeScript durch einen HTTP GET Aufruf erfolgt, wurde die Methode asynchron implementiert. Nachdem der Service die Datei eingelesen und umgewandelt hat, gibt dieser ein *ConfigObject* zurück. Dabei handelt es sich um ein individuelles TypeScript *interface*, welches implementiert wurde, um die Struktur und Einträge der Konfiguration widerzuspiegeln. Durch die Ausgliederung der Konfiguration in eine Datei kann diese im laufenden Betrieb geändert werden, ohne dass es einer Änderung am eigentlichen Code der Webanwendung bedarf.

Der *HttpService* dient zur Abstraktion der Kommunikation mit dem Vorhersagemodell (siehe Code 6.2). Er importiert den *ConfigurationService* mittels DI, um Informationen über die Backend-Adressen zu erhalten. Der *HttpService* wird wiederum als Schnittstelle zu den Backend-Funktionen in die *File Analysis*-View injiziert. Er bietet dabei eine Methode zur Klassifikation von Pneumonien und eine zur Unterscheidung von viralen und bakteriellen Pneumonien. Um sicherzustellen, dass die Benutzeroberfläche bis zur Antwort des Servers reaktionsfähig bleibt, wurden die Methoden asynchron implementiert. Sollten dem Nutzer keine Röntgenbilder zur Verfügung stehen, können

```
1 /**
2 * Sends a http post request for stage1 (lung segmentation) to backend
3 *
4 * @param file x-ray image
5 * @returns an observable containing the cutout
6 */
7 lungSegmentation(file: File): Observable<ArrayBuffer> {
8     ...
9 }
10
11 /**
12 * Sends a http post request for stage2 (pneumonia classification) to
13 * backend
14 * @param file x-ray image
15 * @returns an observable containing the classification results
16 */
17 pneumoniaClassification(file: File): Observable<Object> {
18     ...
19 }
20
21 /**
22 * Sends a http post request for stage3 (viral classification) to backend
23 *
24 * @param file x-ray image
25 * @returns an observable containing the classification results
26 */
27 viralClassification(file: File): Observable<Object> {
28     ...
29 }
```

Code 6.2: Zusammenfassung der Methoden des *HttpService* zur Kommunikation mit dem Backend

Beispielbilder verwendet werden, die als Assets gemeinsam mit der Konfigurationsdatei bereitgestellt werden. Weitere Details zur Schnittstelle zwischen der grafischen Benutzeroberfläche und dem Backend werden in Kapitel 7 diskutiert.

Bevor die Röntgenbilder an den Server gesendet, oder auf der Webseite angezeigt werden können, sind diverse Konvertierungen nötig. Da das Umformen von Daten ein weiterer klassischer Anwendungsfall für die Service-basierte Architektur von Angular ist, wurden diese Umformungen im *ConversionService* gekapselt. Die Übersicht der zur Verfügung gestellten Methoden ist in Code 6.3 zu sehen.

```

1 /**
2 * Converts an ArrayBuffer of an image into a base64 formatted url string
3 *
4 * @param arrayBuffer image to be converted
5 * @returns the base64 encoded string representing the image
6 */
7 arrayBufferToUrlString(arrayBuffer: ArrayBuffer): string {
8     ...
9 }
10
11 /**
12 * Converts a File into a SafeUrl
13 *
14 * @param file image to be converted
15 * @returns a SafeUrl that represents the image
16 */
17 fileToString(file: File): SafeUrl {
18     ...
19 }
20
21 /**
22 * Converts an ArrayBuffer of an image into a File
23 *
24 * @param buffer data to be converted
25 * @param name (optional) filename of the returned File, default: "mask.
26 *               jpg"
27 * @param typeString (optional) MIME type of the returned File, default:
28 *               "image/jpg"
29 * @returns a File object containing the buffer, name and typeString
30 */
31 arrayBufferToFile(buffer: ArrayBuffer, name: string="mask.jpg",
32                   typeString: string="image/jpg"): File {
33     ...
34 }
```

Code 6.3: Übersicht der verfügbaren Datenumformungen im *ConversionService*

6.3.2 Startseite und Tutorial

Wie im vorherigen Abschnitt bereits erwähnt wurde, stellen die Startseite (*Welcome-View*) und das Tutorial (*Tutorial-View*) statische Informationen zur Verfügung. Die Startseite soll dabei eine kurze Übersicht über die Webanwendung geben und verlinkt die Tutorial Seite direkt unter der Überschrift, um Missverständnissen vorzubeugen (siehe Abbildung 6.2). Um den Bestimmungen der verwendeten Daten gerecht zu werden, wird erneut deutlich gemacht, dass es sich hierbei um eine prototypische Anwendung handelt, die nicht für kommerzielle oder medizinische Zwecke verwendet werden darf.

Die Tutorial Seite geht anschließend genauer auf die Nutzung der Seite und Datenschutzaspekte ein. Falls die Seite zur Analyse von Röntgenbildern für manche Nutzer nicht intuitiv bedienbar sein sollte, bietet die Tutorial Seite zudem eine animierte Grafik, welche die Bedienung demonstriert. Die Erläuterungen der Start- und Tutorial-

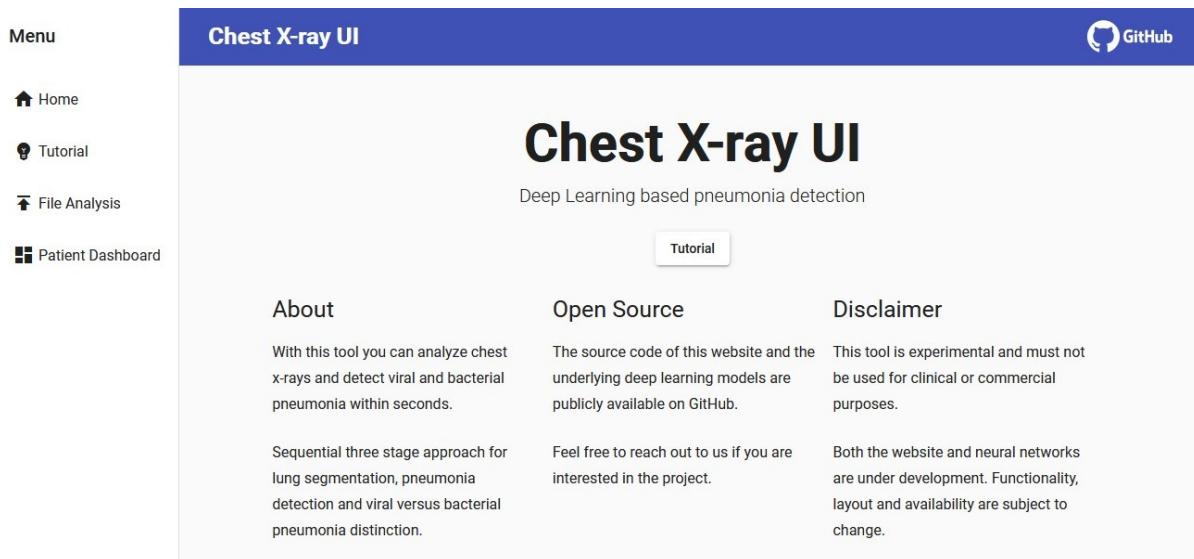


Abbildung 6.2: Startseite der Webanwendung mit allgemeinen Informationen über das Projekt

Seite tragen somit maßgeblich dazu bei, ein möglichst selbsterklärendes System zu entwickeln, wie es in Abschnitt 6.1.2 gefordert wurde.

Die Webanwendung ist grundsätzlich nicht öffentlich zugänglich, sondern ist nur durch eine temporäre Adresse mit Passwortschutz erreichbar. Daher konnte auf die Erstellung eines offiziellen Impressums und einer Seite für Datenschutzrichtlinien verzichtet werden, wie dies bei öffentlichen Webseiten laut Datenschutz-Grundverordnung (DSGVO) notwendig wäre [18].

6.3.3 Analyse von Röntgenbildern

Den Kern der Anwendung stellt die *File Analysis*-View dar. Sie bietet eine Schnittstelle zum Vorhersagemodell für die Analyse von Thoraxröntgenbildern. Wenn der Nutzer die Schaltfläche mit dem Pluszeichen betätigt, öffnet sich ein Dialog zur Dateiauswahl (siehe Abbildung 6.3). Abhängig vom Betriebssystem des Endgeräts kann dieser Dialog gegebenenfalls anders aussehen als in der Abbildung. Der Anwender kann ein Röntgenbild aus seinem System hochladen, welches anschließend zur Analyse an das Backend gesendet wird. Sollte kein Bild zur Verfügung stehen, können durch die Demonstration-Schaltfläche zwei in der Webseite hinterlegte Beispielbilder hochgeladen werden.

Wie in Abschnitt 6.3.1 bereits erläutert wurde, sind die Methoden des *HttpService* zur Kommunikation mit dem Backend asynchron implementiert. Daher ist die Webanwendung auch vor dem Eintreffen der Antwort weiter reaktionsfähig. Um dem Nutzer zu signalisieren, dass die abgesendete Anfrage verarbeitet wird und in Kürze eine Antwort eintrifft, wird eine Ladeanimation angezeigt (siehe Abbildung 6.4).

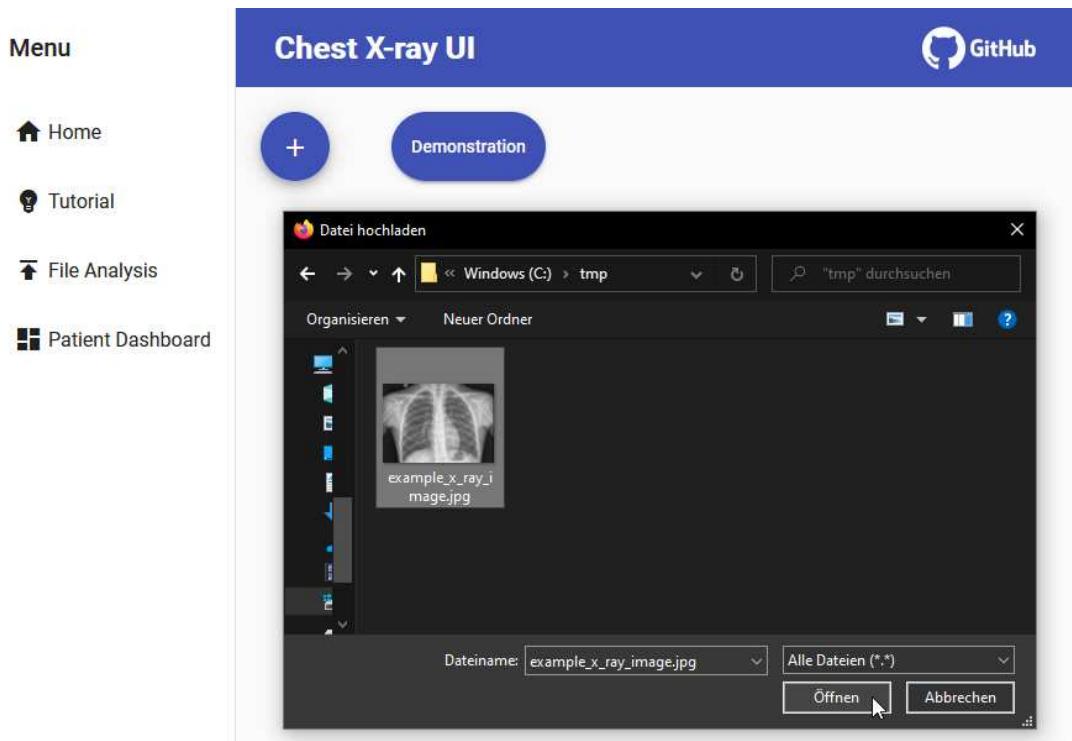


Abbildung 6.3: Dialog zur Auswahl der Datei, die zur Analyse an das Backend gesendet werden soll

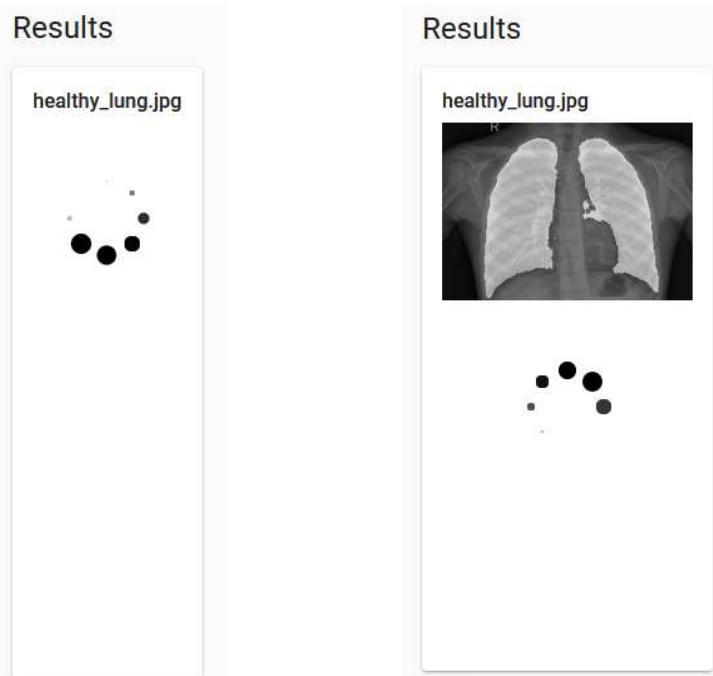


Abbildung 6.4: Ladeanimation zur Überbrückung der Wartezeit, bis die Serverantwort ankommt

Das Ergebnis der Klassifikation ist die vorhergesagte Klasse, welche Werte zwischen 0 und 1 einnehmen kann. Der Wert 0 entspricht dabei der Klasse für eine gesunde Lunge und der Wert 1 einem Patienten mit Pneumonie. Bei der zweiten Klassifikation, der Unterscheidung zwischen bakterieller und viraler Pneumonie, entspricht 0 der bakteriellen und 1 der viralen Pneumonie. Diese vorhergesagten Klassen werden in Form einer Skala dargestellt, damit dem Nutzer bewusst ist, wie deutlich die Entscheidung ausfiel. Das Bild wird den beiden Vorhersagemodellen parallel zur Klassifikation übergeben, um die Antwortzeit zu reduzieren. Sollte sich die erste Klassifikation im grün dargestellten Bereich zwischen 0 und 0,25 befinden, so wird das Ergebnis als sicher genug interpretiert und die zweite Klassifikation als irrelevant angesehen. In diesem Fall wird das Ergebnis der zweiten Klassifikation verworfen und keine zweite Skala angezeigt.

Der vom U-Net (siehe Abschnitt 1.1) erkannte Lungenbereich wird halbtransparent über dem Röntgenbild dargestellt. Wenn der Nutzer mit der Maus über das Bild fährt, verschwindet diese Einblendung und das unveränderte Röntgenbild ist zu sehen. Diese Funktion ist in Abbildung 6.5 zu sehen, wobei im linken Bild das unveränderte Röntgenbild zu sehen ist und im rechten Bild das Ergebnis der Lungensegmentierung überlagert dargestellt wird. Wie in Abschnitt 5.3.1 beschrieben wurde, konnte die Segmentierung der Lungenregionen keine Verbesserung der Vorhersagen erreichen. In Zukunft könnte diese Vorgehensweise verwendet werden, um anstelle der Lungenregionen die Heatmap der Aktivierungsfunktionen des Modells überlagert darzustellen. So könnte der Nutzer zusätzlich zu den Skalen der Klassifikationsergebnisse Informationen über die Regionen erhalten, die für die Entscheidung von besonderem Interesse waren.

Die Ergebnisse werden dabei in Kacheln dargestellt, die eine einheitliche Größe besitzen. Wenn mehrere Bilder nacheinander hochgeladen werden, werden die Ergebnisse zunächst in mehreren Kacheln nebeneinander dargestellt. Wird der Bildschirmrand erreicht, werden die nächsten Ergebnisse automatisch in einer neuen Zeile dargestellt. Sollte auch der untere Bildschirmrand erreicht werden, wird die View vertikal scrollbar, wobei die Navigationsleiste und die Titelleiste auch beim Scrollen an ihrer Position fixiert bleiben.

Eine Persistenz der Ergebnisse und Zustände der Anwendung ist bewusst nicht vorgesehen, da dies für Demonstrationszwecke irrelevant ist. Weiterhin wäre eine dauerhafte Speicherung aus Sicht des Datenschutzes problematisch. Sollte das System dennoch in diesem Aspekt erweitert werden, könnte für den prototypischen Einsatz beispielsweise die Web Storage API verwendet werden, welche es unkompliziert ermöglicht, Schlüssel-Wert-Paare im Speicher des Browsers abzulegen [49]. Selbstverständlich wäre jedoch der Einsatz einer vollwertigen Datenbank für echte Persistenz oder gar eine produktive Nutzung des Systems unabdingbar.

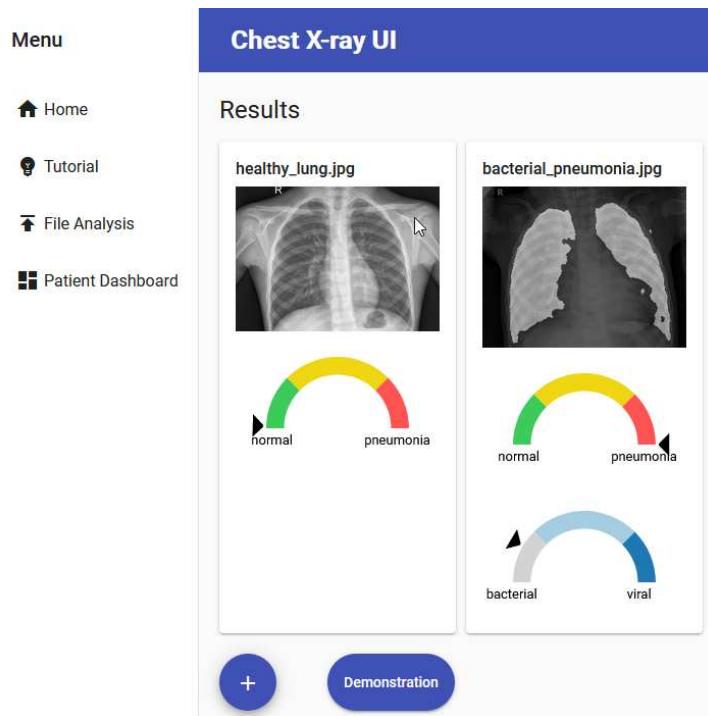


Abbildung 6.5: Ergebnisse der Klassifikation der beiden hinterlegten Beispielbilder zur Demonstration der Röntgenbildanalyse

6.3.4 Patientenübersicht

Durch die modulare Architektur der Anwendung ist es besonders einfach Erweiterungen umzusetzen. Um diese Erweiterbarkeit zu demonstrieren und zu zeigen, dass der Funktionsumfang nicht auf die Analyse von Röntgenbildern beschränkt sein muss, wurde eine weitere Seite erstellt. Die Patientenübersicht stellt Patienten als Kacheln dar und soll zeigen, dass die Verwaltung von personenbezogenen Daten ohne großen Aufwand in die Anwendung integriert werden könnte. Somit konnte die in Abschnitt 6.1.5 beschriebene Anforderung zur Demonstration der Erweiterbarkeit erfüllt werden.

Abbildung 6.6 zeigt Patientenübersicht. Wenn der Nutzer auf das Plussymbol klickt, wird von der *Patient Dashboard*-View die *AddPatientDialog*-Component aufgerufen und eine Eingabemaske erscheint im Vordergrund der Anwendung. Da die Patientenübersicht nur als Demonstration der Erweiterbarkeit gedacht ist, wird in der Eingabeaufforderung lediglich nach dem Namen des Patienten gefragt.

Nachdem der Nutzer die Eingabe bestätigt hat, erscheint eine neue Kachel mit dem eingegebenen Namen und einem kurzen Informationstext über denkbare Inhalte, die in die Kachel integriert werden könnten (siehe Abbildung 6.6). In der rechten oberen Ecke der Kachel befindet sich zudem eine Schaltfläche, die eine Liste weiterer Aktionen zur Verfügung stellt. Das Icon der Schaltfläche und die Liste sind dabei Elemente aus dem Material Design Framework, um eine intuitive und bereits vertraute Nutzung zu ermöglichen. Als Aktionen der Auswahlliste stehen das Hinzufügen von Röntgenbildern und das Entfernen des Patienten zur Verfügung. Beim Hinzufügen

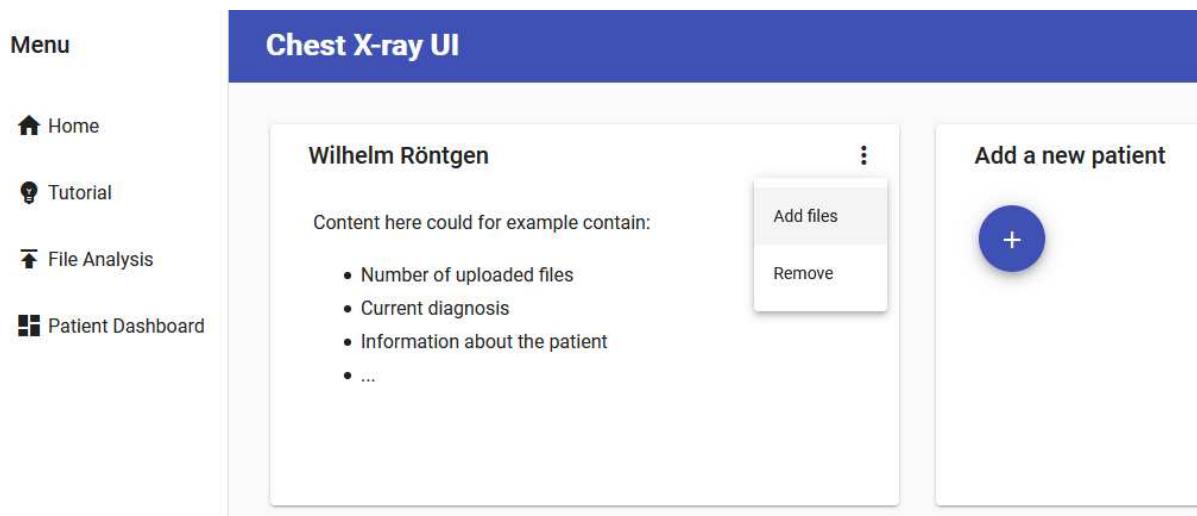


Abbildung 6.6: Patienten erscheinen in der Übersicht als Kachel. Über das Kontextmenü jeder Kachel können Aktionen ausgeführt werden.

eines Röntgenbildes erscheint ebenso wie im vorherigen Abschnitt ein Dialog zur Auswahl einer Datei. Nach der Auswahl der Datei wird diese jedoch bewusst verworfen, da die Patientenübersicht nur für Demonstrationszwecke gedacht ist und nicht mit dem Backend verbunden ist (siehe Architektschaubild in Abbildung 6.1). Wenn die Schaltfläche zum Entfernen eines Patienten ausgewählt wird, verschwindet dieser von der Patientenübersicht.

6.3.5 Mobile Ansicht

Wie bereits in Abschnitt 6.1.4 definiert wurde, sollte die Anwendung möglichst unabhängig vom Endgerät einsetzbar sein. In Abbildung 6.7 sind die Views *Welcome* und *File Analysis*, sowie die Navigationsleiste in der mobilen Ansicht zu sehen. Mobile Geräte werden in der Regel im Hochkant-Format verwendet, weshalb die Navigation in der Anwendung möglichst vertikal stattfinden sollte. Daher werden bei Verwendung eines mobilen Geräts mehrspaltige Elemente der Views in einspaltige Elemente umgewandelt. Durch den Wechsel in einen einspaltigen Modus erfolgt die Navigation selbst auf kleinen Bildschirmen ausschließlich vertikal. Zudem sorgt diese Gestaltung dafür, dass das Erscheinungsbild der Anwendung einer nativen App ähnelt.

Um Platz einzusparen, wird die Navigationsleiste in der mobilen Ansicht eingeklappt. Die Navigationsleiste ist dann durch das aus dem Material Design bekannte Menü-Icon im linken oberen Bildbereich ausklappbar. Ebenso wie der zuvor erwähnte einspaltige Modus, ist die Darstellung und Nutzung der Navigationsleiste an native Apps angelehnt, um eine intuitive Bedienung auf mobilen Endgeräten zu ermöglichen. Bei einem kleinen bis mittelgroßen Tablet wird die gleiche Darstellung wie bei Smartphones genutzt, während bei besonders großen Tablets die Desktopansicht verwendet wird.

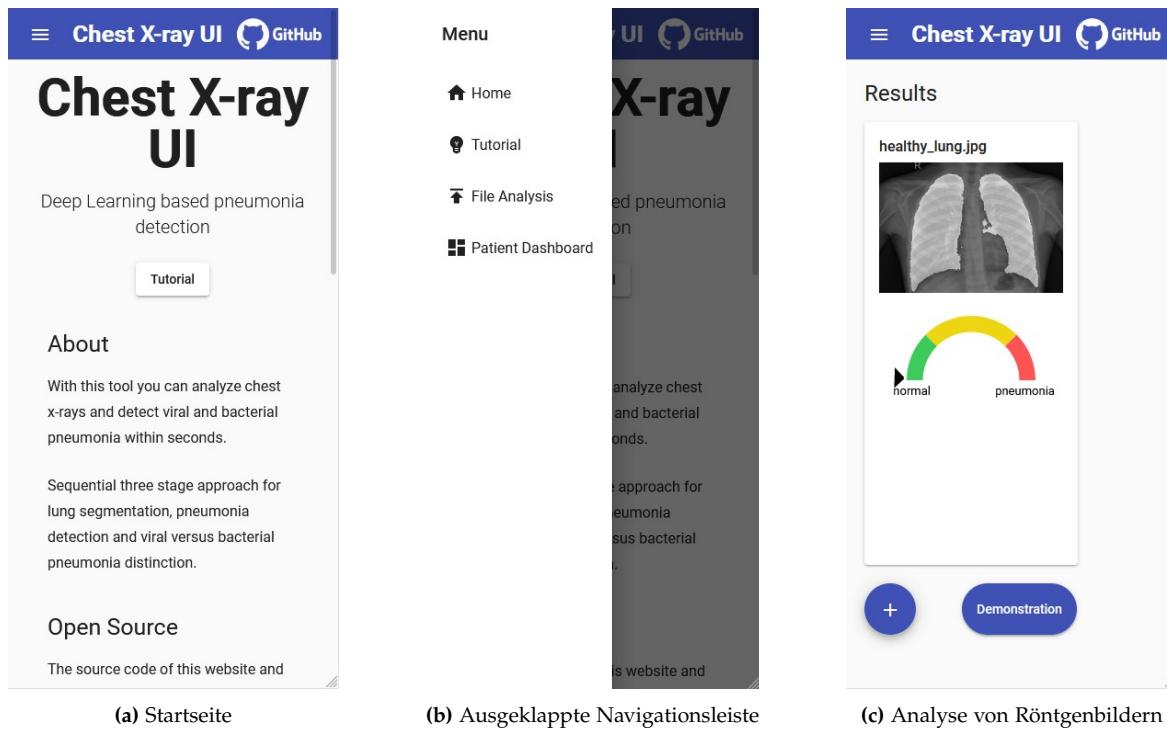


Abbildung 6.7: Beispiele für das Erscheinungsbild der Webanwendung beim Verwenden eines mobilen Endgeräts

6.4 Vergleich mit anderen Ansätzen

Als Abschluss dieses Kapitels wird ein kurzer Vergleich mit anderen Ansätzen erläutert und die jeweiligen Vor- und Nachteile beleuchtet.

6.4.1 Native Anwendungen

Anwendungen werden als **nativ** bezeichnet, wenn sie speziell für ein bestimmtes Betriebssystem entwickelt wurden. Native Anwendungen müssen in der Regel als Software auf dem Gerät installiert werden. Bei mobilen Endgeräten kann dieser Installationsvorgang gegebenenfalls durch die Nutzung eines App-Stores erfolgen. Im Vergleich zur Webanwendung bieten native Applikationen eine höhere Performanz und Kompatibilität zur Hardware. Bei sehr rechenintensiven und komplexen Applikationen kann die Beschränkung auf ein Betriebssystem daher durchaus sinnvoll sein.

Wenn jedoch eine breite Zielgruppe mit unterschiedlichen Endgeräten angesprochen werden soll, sind native Anwendungen eher schlecht geeignet. Pro Betriebssystem wäre es notwendig, eine spezielle Version der Anwendung für dieses System zu entwickeln. Da für die Umsetzung der Anwendung in dieser Arbeit möglichst viel Flexibilität erwünscht ist, würde diese Tatsache somit ein großes Problem darstellen.

Weiterhin ist eine Aktualisierung der Installation notwendig, wenn eine neue Version der nativen Anwendung zur Verfügung steht. Im Gegensatz dazu kann die hier

entwickelte Webanwendung im laufenden Betrieb verändert werden und der Nutzer erhält automatisch beim nächsten Ladevorgang der Seite alle Neuerungen.

Auch wenn native Anwendungen einige Vorteile bieten, überwiegen für die Anwendungsfälle dieser Arbeit die negativen Aspekte. Eine hohe Rechenintensität innerhalb der Anwendung wird nicht benötigt, da die aufwändigen Berechnungsschritte zur Analyse der Röntgenbilder im Backend ausgeführt werden. Des Weiteren ermöglichen es moderne Frameworks, wie das hier verwendete Angular Framework, dem Nutzer einer Webanwendung ein annähernd natives Erlebnis zu bieten. Besonders im Bereich der Webanwendungen, die für mobile Endgeräte geeignet sind, konnten in den vergangenen Jahren große Fortschritte erzielt werden.

6.4.2 Lokale Klassifikation

Alternativ zum Übertragen der Röntgenbilder an einen Server, kann die Klassifikation auch direkt im Browser des Endgeräts vorgenommen werden. Für diesen Ansatz müssen JavaScript Bibliotheken für Deep Learning wie zum Beispiel TensorFlow.js genutzt werden [1]. Das Vorhersagemodell wird beim Aufrufen der Seite an den Browser des Geräts gesendet und dort lokal ausgeführt. Diese Herangehensweise wurde von Cohen u. a. bei der Entwicklung der Anwendung Chester verwendet [13].

Vorteilhaft an diesem Ansatz ist, dass keinerlei Kommunikation medizinischer Daten mit dem Server stattfindet. Somit können Bedenken hinsichtlich des Datenschutzes eliminiert werden und es ist sichergestellt, dass serverseitig keine Weiterverarbeitung oder Speicherung von persönlichen Daten stattfindet. Weiterhin kann die Webanwendung offline betrieben werden, sobald das Vorhersagemodell fertig geladen wurde. Das Laden des Modells und somit auch das Laden der gesamten Anwendung kann jedoch relativ lange dauern, da tiefe neuronale Netze mit vortrainierten Gewichten viel Speicherplatz benötigen.

Wie bereits im vorherigen Abschnitt erklärt wurde, sollten rechenintensive Anwendungen eher als native Applikationen umgesetzt werden. Da Klassifikation der Röntgenbilder aufwändig ist, muss bei Verwendung dieses Ansatzes zwangsläufig mit Einbußen der Performance gerechnet werden. Bei der Verwendung der Chester Anwendung zeigt sich zudem, dass besonders bei mobilen Endgeräten die Klassifikation erheblich länger dauert als mit einem Desktop oder Laptop. Hier gerät die Hardware in Kombination mit der Webanwendung eindeutig an ihre Grenzen. Da die Integration von Deep Learning-Technologien in Browser noch jung und nicht sehr verbreitet ist, kann es bei älteren Geräten außerdem zu Kompatibilitätsproblemen kommen.

Kapitel 7

Bereitstellung des Systems

In den vorherigen Kapiteln wurde das Vorhersagemodell und die grafische Benutzeroberfläche erläutert. Um ein gemeinsames Agieren dieser zu ermöglichen, wird in diesem Kapitel eine Strategie zur Bereitstellung und Vernetzung der Komponenten erarbeitet. Ähnlich wie bei der Entwicklung der Webanwendung, liegt hierbei ein besonderes Augenmerk auf Flexibilität und Modularität.

Im ersten Abschnitt werden die verwendeten Amazon Web Services (AWS) Dienste erläutert. Anschließend wird die Architektur und das Zusammenspiel der Komponenten näher erläutert. Da bei der Verwendung von Cloud-Technologien häufig Bedenken zum Datenschutz und der Privatsphäre vorherrschen, wird zum Abschluss des Kapitels dieser Aspekt im Hinblick auf das entwickelte System diskutiert.

7.1 Verwendete Cloud-Computing-Dienste

AWS ist der international führende Cloud-Computing-Anbieter und stellt eine Vielzahl verschiedener Dienste zur Verfügung [3]. Berichten zufolge nutzten im Jahr 2019 über 80 % der im DAX gelisteten Konzerne AWS [7]. Die Dienste sind dabei so gestaltet, dass sie effektiv über Application Programming Interfaces (APIs) und Software Development Kits (SDKs) angesprochen werden können. Für manuelle Interaktion kann die Benutzerschnittstelle AWS-Managementkonsole verwendet werden. Cloud-Computing bietet den Vorteil, dass Server-Ressourcen dynamisch skaliert werden können. Somit kann spontan auf hohe Auslastungen reagiert werden und bei geringer Nachfrage oder gar Leerlauf die Kosten optimiert werden. Auf der negativen Seite ist anzumerken, dass Anwendungen teilweise auf den Cloud-Anbieter zugeschnitten sein müssen, da die APIs und SDKs herstellerspezifisch sind.

7.1.1 Simple Storage Service

Amazon Simple Storage Service (S3) ist ein Objektspeicherservice, der verschiedene Konfigurationen für Skalierbarkeit, Datenverfügbarkeit, Sicherheit und Leistung bietet. Die abgelegten Daten können jeglicher Art sein und werden in Buckets und Objects

organisiert, welche konzeptionell vergleichbar mit Verzeichnissen und Dateien sind. Abgesehen von der reinen Datenspeicherung kann der Dienst auch für das Hosting statischer Webseiten verwendet werden. Eine Webseite wird in diesem Kontext als statisch bezeichnet, wenn ihr Aufbau unabhängig von serverseitiger Verarbeitung ist, wie es beispielsweise bei PHP: Hypertext Preprocessor (PHP) der Fall ist. Die Inhalte statischer Seiten können dennoch sehr interaktiv sein und viel Logik beinhalten, da auf Seite des Client Sprachen wie JavaScript und TypeScript zur Verfügung stehen.

7.1.2 CloudFront

Der Dienst Amazon CloudFront ist ein sogenanntes Content Delivery Network (CDN) und hat das Ziel, dem Benutzer eine höhere Verfügbarkeit und schnellere Ladezeiten zu ermöglichen. Realisiert werden diese Vorteile durch ein verteiltes Netz aus Proxy-Servern und Datenzentren, welche möglichst nah am Endbenutzer lokalisiert sind, um Latenzzeiten zu reduzieren. Schnellere Ladezeiten und Datenübertragungen werden beispielsweise dadurch erreicht, dass datenintensive Inhalte wie Medien auf den verteilten Servern zwischengespeichert werden. Abgesehen vom besseren Benutzerlebnis bietet CloudFront die Möglichkeit, als Zugangspunkt für dahinterliegende AWS Dienste zu agieren. Dadurch erhält die Anwendung eine zusätzliche Abstraktionsschicht, was die Vernetzung der Dienste erleichtert und architektonische Vorteile mit sich bringt.

7.1.3 Certificate Manager

Der Certificate Manager ist ein Dienst, mit dem digitale Zertifikate, wie zum Beispiel Transport Layer Security (TLS)-Zertifikate, verwaltet werden können. Digitale Zertifikate werden verwendet, um die Netzwerkkommunikation zu verschlüsseln und die Authentizität und Integrität von Webseiten im Internet nachzuweisen. Mit dem Certificate Manager können Zertifikate schnell und unkompliziert für andere AWS-Dienste angefordert und bereitgestellt werden. Zusätzlich hilft der Dienst dabei, die Lebenszyklen der Zertifikate zu verwalten und gegebenenfalls Zertifikaterneuerungen anzustoßen.

7.1.4 Elastic Load Balancing

Damit aufwändige Berechnungen oder eine hohe Anzahl von Anfragen effizient bearbeitet werden, kann der Gesamtaufwand durch Lastverteilung (englisch: *Load Balancing*) auf mehrere parallel arbeitende Systeme verteilt werden. Elastic Load Balancing (ELB) stellt vier anwendungsspezifische Load Balancer in Form von AWS-Diensten zur Verfügung.

- Application Load Balancer (HTTP/HTTPS)

- Network Load Balancer (TCP/UDP)
- Gateway Load Balancer (Firewalls)
- Classic Load Balancer (Sonstiges)

Der eintreffende Datenverkehr wird durch die Load Balancer beispielsweise zu Elastic Compute Cloud (EC2)-Instanzen oder Docker-Containern weitergeleitet. Wenn sich die benötigte Rechenleistung mit der Zeit ändert, können die Leistung und die Kosten durch das Zu- und Abschalten von Instanzen dynamisch skaliert werden. Nachdem sich die Infrastruktur durch die automatische Skalierung verändert hat, werden die Regeln zur Weiterleitung der Daten ebenfalls entsprechend geändert. Der Load Balancer steht für Anwendungen dabei als zentraler Kommunikationspartner und Schnittstelle zur Verfügung, sodass die dahinterliegenden dynamischen Strukturen verborgen bleiben.

7.1.5 Elastic Container Registry

Elastic Container Registry (ECR) ist ein Managed Service zum Speichern und Verwalten von Docker-Images. Unter einem Managed Service versteht man einen Dienst, bei dem sowohl die Hardware als auch die Software vom Anbieter verwaltet und betreut werden [vgl. 39, S. 340]. Der Anwender überträgt somit die Verantwortung für die Verfügbarkeit, Wartung und Leistungsfähigkeit des Service an AWS. In ECR können Lebenszyklen für Images definiert werden, die anhand von festgelegten Regeln alte Versionen aus dem System entfernen. Weiterhin bietet der Dienst die Möglichkeit, dass Images automatisch nach möglichen Schwachstellen oder enthaltener Schadsoftware durchsucht werden.

7.1.6 Elastic Container Service

Elastic Container Service (ECS) ist ein Dienst zum Ausführen, Anhalten und Verwalten von Docker-Containern in einem Cluster. Abhängig von den Bedürfnissen des Anwenders kann er ein Cluster aus EC2-Instanzen selbst verwalten, oder den Managed Service AWS Fargate in Anspruch nehmen. Welche Ressourcen, Images und weitere Eigenschaften einem Container zugewiesen werden, wird in Tasks definiert. Eine oder mehrere Tasks können dabei zu einem Service zusammengefasst werden, welcher auf dem Cluster verteilt wird. ECS bietet die Möglichkeit die Rechenkapazitäten des Clusters dynamisch zu skalieren. Sollte mehr Rechenleistung benötigt werden, können weitere EC2-Instanzen automatisch gestartet werden. ECS kommuniziert zudem mit der ECR und erkennt, wenn sich die Versionen der Docker-Images ändern. Sollte ein neuer Container bereitgestellt, oder ein bestehender neu gestartet werden, kümmert sich der Dienst automatisch um die Aktualisierung der Container auf die neueste Version des Images.

7.2 Umsetzung

In diesem Abschnitt wird beschrieben, wie das Vorhersagemodell und die Webanwendung, durch die im vorherigen Abschnitt beschriebenen AWS Dienste, bereitgestellt und vernetzt werden.

7.2.1 Backend

Bevor die Bestandteile des Systems in der Cloud betrieben werden können, müssen sie auf die jeweiligen Dienste zugeschnitten werden. Die Funktionalität des Vorhersagemodells wird dabei als Webserver zur Verfügung gestellt. Durch eine Representational State Transfer (REST) Schnittstelle können Bilder an das neuronale Netz übermittelt werden. Diese werden klassifiziert und das Ergebnis anschließend als Antwort des Aufrufs zurückgesendet. Zur Implementierung der Endpunkte wurde das Python Framework Flask verwendet [80]. In Code 7.1 ist der Minimalaufbau eines solchen Webservers dargestellt. Um das Beispiel kompakt und verständlich zu halten, ist nur ein Teil der gesamten Endpunkt-Implementierung dargestellt.

```

1 from tensorflow.keras.models import load_model
2 from flask import Flask
3
4 app = Flask(__name__)
5 model = load_model("<model_name>.h5", compile=False)
6
7 # Route that offers prediction functionality as REST call
8 @app.route('/predict', methods=['POST'])
9 def predict_pneumonia():
10     # Data extraction, preprocessing, etc.
11     ...
12     # Classification
13     prediction = model.predict(...)
14     # Postprocessing the result
15     ...
16     # Return classification results
17     return {
18         "result": result,
19         "prediction_value": prediction_value,
20     }
21
22 if __name__ == '__main__':
23     app.run()    # Start webserver

```

Code 7.1: Beispielhafte Struktur eines Endpunkts in Python mit einer Route zur Kommunikation mit dem neuronalen Netz

Jedes Vorhersagemodell wird in einem eigenen Webserver gekapselt und als eine Art Microservice zur Verfügung gestellt. Somit sind die Modelle vollständig voneinander entkoppelt, was die Wartbarkeit der Netze erleichtert. Sollten Änderungen an einem der Modelle vorgenommen werden, ist es ausreichend den betroffenen Webserver zu

aktualisieren, ohne die verbleibenden zu beeinflussen. Zudem kann durch die Kapselfung das Potential von ECS genutzt werden, um bei Bedarf zusätzliche Rechenleistung für einen der Microservices zur Verfügung zu stellen.

Damit die Webserver von ECS als Docker-Container bereitgestellt werden können, ist es zuvor jedoch notwendig, diese als Docker-Images in ECR hochzuladen. Deshalb wurde für jeden Microservice ein Dockerfile implementiert (siehe Code 7.2), welches die benötigten Inhalte sammelt und anschließend einen gunicorn Server konfiguriert. Für die Bereitstellung der in Flask implementierten Endpunkte werden externe Server wie gunicorn oder Nginx verwendet, da der mitgelieferte Server von Flask nur zum Entwickeln und Testen gedacht und daher nicht ausreichend optimiert und performant genug ist. Nachdem die Docker-Images in ECR hochgeladen sind, kann die Bereitstellung des Backend durchgeführt werden.

```
1 # Use Python image as base
2 FROM python:3.7-slim
3 # Create folders and install required python packages
4 RUN mkdir /app
5 WORKDIR /app
6 RUN mkdir ./model
7 COPY requirements.txt .
8 RUN pip3 install -r requirements.txt
9 # Create user and copy files
10 RUN groupadd appgroup
11 RUN useradd -ms /bin/bash -d /app -G appgroup appuser
12 USER appuser
13 COPY app.py .
14 COPY model/* model/
15 COPY modules/* modules/
16 # Configure gunicorn Webserver
17 CMD gunicorn --bind 0.0.0.0:5000 app:app
```

Code 7.2: Dockerfile zur Erstellung eines Webserver Image

ECS wurde so konfiguriert, dass die Vorhersagemodelle immer das aktuellste Image verwenden, welche durch den Tag `:latest` markiert sind. Wenn ein Webserver hochgefahren wird, startet er somit immer in der aktuellen Version. Da diese Arbeit einen Prototyp entwickelt, ist das automatische Verwenden der aktuellen Version unproblematisch. In einer produktiv eingesetzten Anwendung wären der Test- und Bereitstellungsprozess selbstverständlich umfangreicher. Wird nun eine neue Version des Images verfügbar, kann in der gleichen Instanz parallel ein zweiter Container gestartet werden. Der neue Container wird die aktualisierte Version des Images, und somit auch die aktualisierte Version des Vorhersagemodells, enthalten. Wenn der Container fertig hochgefahren ist, kann der vorherige Container entfernt werden. Somit ist der Übergang zu einem neuen Modell ohne Ausfallzeit des Webservers mit minimalem Aufwand möglich.

7.2.2 Frontend

Wie in Abschnitt 7.1.1 bereits beschrieben wurde, kann ein S3 Bucket für das Hosting statischer Webseiten verwendet werden. Da Angular Anwendungen statische Webseiten sind, ist die Bereitstellung der Webanwendung in S3 ohne Probleme möglich. Die Anwendung muss lediglich kompiliert werden, ein Hauptdokument mit dem Namen „index.html“ zur Verfügung stellen und in den S3 Bucket kopiert werden. Für die Kompilierung der Applikation wird der Befehl `ng build --prod` verwendet. Die Option `--prod` aktiviert den Produktivmodus und mit ihm diverse Optimierungen, welche die Performance der Anwendung verbessern.

Um die Bereitstellung neuer Funktionalitäten in der Webanwendung zu beschleunigen, wurden Methodiken des Continuous Deployment angewandt. Der Begriff Continuous Deployment beschreibt eine Veröffentlichungsstrategie für Software, bei der die Inbetriebnahme einer neuen Softwareversion möglichst automatisiert durchgeführt wird. Durch die Automatismen können Änderungen schneller in das Produktivsystem integriert und redundanter Aufwand vermieden werden.

Umgesetzt wurde dieser Continuous Deployment Gedanke durch GitHub Actions. In GitHub Actions kann durch die Verwendung sogenannter Workflows Prozesse und Abläufe definiert werden, die durch Aktivitäten im Repository angestoßen werden. Workflows werden dabei in der Sprache YAML Ain't Markup Language (YAML) geschrieben, welche durch Einrückungen strukturiert ist und Daten ähnlich wie JSON in Key-Value-Paaren organisiert.

In Code 7.3 ist der Workflow zur Bereitstellung der Webanwendung zu sehen, wobei einige weniger relevante Bereiche bewusst nicht enthalten sind. Zu Beginn wird der Name des Workflows festgelegt und Auslöser definiert, für die der Workflow angestoßen werden soll. Der Workflow wurde so konfiguriert, dass er angestoßen wird, wenn auf dem master Branch Änderungen im Ordner der Webanwendung vorgenommen werden. Die Ausführung des Workflows geschieht auf virtuellen Maschinen, die von GitHub betrieben werden, wobei in dieser Arbeit eine Ubuntu Instanz verwendet wurde. Auf dieser virtuellen Maschine werden dann die konfigurierten Schritte sequenziell ausgeführt.

Bei den hier implementierten Schritten wird zu Beginn das System eingerichtet und benötigte Pakete installiert, um anschließend den Angular Buildprozess mit dem zuvor erwähnten Befehl `ng build --prod` anzustoßen. Als nächstes wird die JSON-Konfigurationsdatei durch ein Bash-Skript konstruiert und abgespeichert. GitHub Actions stellt Anbindungen an viele AWS Dienste zur Verfügung. Daher kann im letzten Schritt des Workflows die gesamte Webanwendung in den S3 Bucket hochgeladen werden. Damit Adressen, IDs und Passwörter nicht als Klartext im Workflow zu sehen sind, wurden diese als verschlüsselte GitHub Secrets eingerichtet. Sollten in einem Schritt Fehler auftreten, wird die Ausführung des Workflows abgebrochen und die fehlerhafte Anwendung nicht in AWS bereitgestellt. Durch die Verwendung

```

1 # This is a workflow for continuous deployment of Chest X-ray UI
2 name: Frontend deployment
3
4 # When to execute the workflow
5 on:
6   push:
7     branches: # At every commit to master branch
8       - master
9     paths:    # But only when changes to frontend are made
10      - 'frontend/**'
11
12 # Workflow to execute
13 jobs:
14   build:
15     runs-on: ubuntu-latest
16
17   steps:
18     # Check-out repository under $GITHUB_WORKSPACE
19     ...
20     # Setup Node.js and install packages
21     ...
22     # Build Angular application
23     - name: Build
24       run: ng build --prod
25       working-directory: ./frontend
26     # Add backend URLs from GitHub secrets to config.json
27     - name: Configure Backend URL
28       run: |
29         JSON_CONFIG='{"backend": {"stage1_url": "%s", "stage2_url": "%s"} }\n'
30         printf "$JSON_CONFIG" "$BACKEND_URL_STAGE1" \
31             "$BACKEND_URL_STAGE2" >> ./frontend/dist/chest-ui/assets/
32             config/config.json
33   shell: bash
34   env:
35     BACKEND_URL_STAGE1: ${{ secrets.BACKEND_URL_STAGE1 }}
36     BACKEND_URL_STAGE2: ${{ secrets.BACKEND_URL_STAGE2 }}
37   # Deploy application to S3 bucket
38   - name: Deploy
39     uses: jakejarvis/s3-sync-action@master
40     with:
41       args: --acl public-read --follow-symlinks --delete --exclude 'assets/config/README.md' --exclude 'assets/config/
42           example_config.json'
43   env:
44     AWS_S3_BUCKET: ${{ secrets.AWS_S3_BUCKET }}
45     AWS_ACCESS_KEY_ID: ${{ secrets.AWS_ACCESS_KEY_ID }}
46     AWS_SECRET_ACCESS_KEY: ${{ secrets.AWS_SECRET_ACCESS_KEY }}
47     AWS_REGION: 'eu-central-1'
48     SOURCE_DIR: './frontend/dist/chest-ui'

```

Code 7.3: GitHub Action zur automatischen Bereitstellung neuer Versionen der Webanwendung

von GitHub Actions konnte somit eine vollständige Continuous Deployment Pipeline implementiert werden, die bei Änderung des Quellcodes die Webanwendung vollautomatisch aktualisiert.

7.2.3 Architektur

Die beiden vorherigen Abschnitte haben beschrieben, wie die Webanwendung und die Vorhersagemodelle in AWS bereitgestellt werden. In diesem Abschnitt wird die Gesamtstruktur des Systems und die Vernetzung der zuvor beschriebenen Komponenten erklärt. Abbildung 7.1 ermöglicht einen Überblick der Bestandteile und wird nachfolgend genauer erläutert.

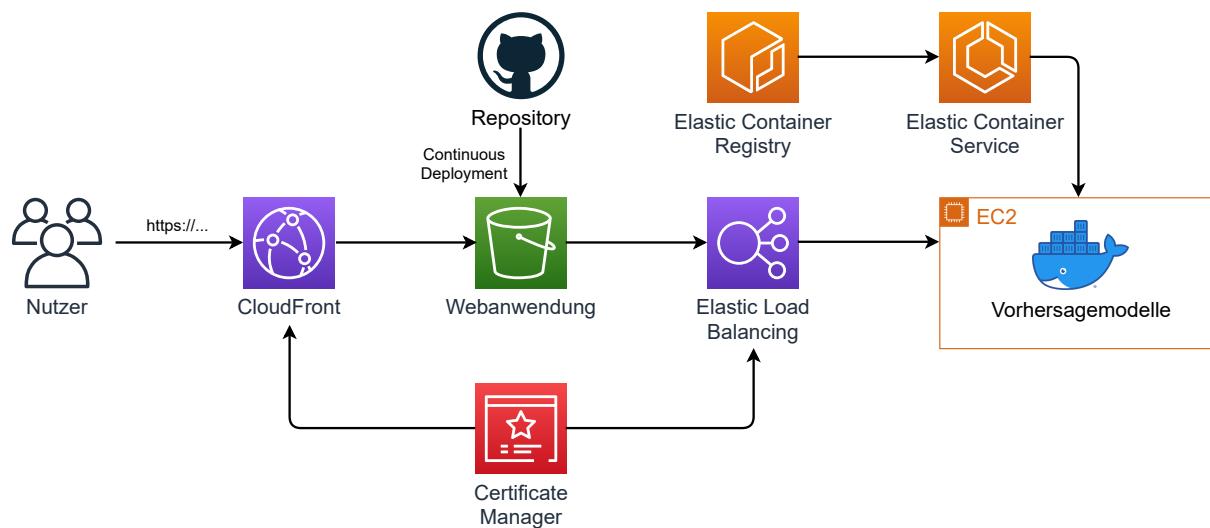


Abbildung 7.1: Architektschaubild der AWS Dienste zur Bereitstellung des Assistenzsystems

Der CDN-Dienst CloudFront dient für den Nutzer als Einstiegspunkt in die Anwendung und stellt die aufzurufende Adresse zur Verfügung. Weiterhin fordert er vom Certificate Manager das TLS-Zertifikat zur Verschlüsselung der Kommunikation an. Anfragen an CloudFront werden anschließend an die Webanwendung im S3 Bucket weitergeleitet. Wie in Abschnitt 7.2.2 und Code 7.3 ausführlich diskutiert wurde, wird der Inhalt des S3 Bucket durch GitHub Actions direkt aus dem Repository kontinuierlich aktualisiert.

Als Schnittstelle zwischen der Webanwendung und den Vorhersagemodellen, wird ein Application Load Balancer (ALB) aus dem Dienst ELB verwendet. Er nimmt REST Anfragen der Benutzeroberfläche entgegen und leitet diese an die jeweiligen Docker-Container weiter. Diese Vorgehensweise bietet den Vorteil, dass die Struktur der EC2-Instanzen und Container abstrahiert und eine Entkopplung der Webanwendung von den Vorhersagemodellen erreicht wird. Der ALB dient somit als zentraler Kommunikationspartner für die Benutzeroberfläche, weshalb die Struktur des ECS-Clusters dynamisch verändert werden kann, ohne dass es einer Änderung der Webanwendung

bedarf. Als Ziele des Load Balancers werden sogenannte Target Groups definiert, welche den ECS Services entsprechen. Da pro Vorhersagemodell ein Service implementiert wurde, hat der ALB drei Target Groups. Für die Weiterleitung der Anfragen wurden Regeln definiert, die anhand der verwendeten Route entscheiden, an welche Target Group die Anfrage gerichtet ist. Damit die Anfrage von den Vorhersagmodellen bearbeitet werden kann, muss diese zuerst entschlüsselt werden. Daher stellt der Certificate Manager dem ALB, wie bereits bei CloudFront, das TLS-Zertifikat zur Verfügung.

Wie bereits in Abschnitt 7.2.1 erläutert wurde, sind die Vorhersagmodelle als Container Anwendungen umgesetzt. Die Images werden in ECR hochgeladen und mit einem Tag versehen, damit sie von anderen Diensten abgerufen werden können. ECR ist dabei kompatibel mit dem Kommandozeilenprogramm von Docker, was die Verwendung bereits bekannter Abläufe möglich macht.

In einem ECS-Cluster können mehrere EC2-Instanzen und Services enthalten sein, wobei jeder Service wiederum mehrere Tasks und jede Task mehrere Container enthalten kann. Der Service ist dafür zuständig, seine Tasks auf dem ECS-Cluster zu starten und sorgt dafür, dass stets die konfigurierte Anzahl von Instanzen pro Task verfügbar sind. Eine Task definiert die Anzahl, Parameter, Rechenleistung und Speicherkapazität, mit der die Container gestartet werden sollen. Für jedes Vorhersagmodell wurde in ECS ein eigener Service, beziehungsweise eine eigene Task angelegt. Für den Anwendungsfall dieser Arbeit ist es ausreichend eine EC2-Instanz und einen Container pro Vorhersagmodell zu betreiben. Eine Ausnahme stellt hierbei die Aktualisierung eines Docker-Images dar. Wie in Abschnitt 7.2.1 beschrieben wurde, ist es dann sinnvoll zwei Container für einen kurzen Zeitraum parallel zu betreiben.

Um die Verfügbarkeit der Services zu garantieren, führt der ALB regelmäßig einen obligatorischen Health Check durch. Bei der Ausführung eines Health Checks wird einer Target Group über eine dedizierte Route eine Anfrage gesendet, auf die der enthaltene Service antworten muss. Im Load Balancer kann konfiguriert werden, in welcher Frequenz die Health Checks durchgeführt werden sollen, in welchem Zeitraum eine Antwort erfolgen muss und nach wie vielen fehlgeschlagenen Überprüfungen eine Target Group als fehlerhaft klassifiziert wird. Die Implementierung der Route zur Beantwortung von Health Checks im Flask Webserver ist in Code 7.4 dargestellt. Sollte ein Service als fehlerhaft angesehen werden, wird der ALB keine eintreffenden Anfragen mehr an diesen Service weiterleiten und kann gegebenenfalls weitere Maßnahmen zur Wiederherstellung der Verfügbarkeit einleiten.

```

1 import json
2
3 # Route that answers the health check of the Application Load Balancer
4 @app.route('/health')
5 def healthcheck():
6     return json.dumps({'success':True}), 200, {'ContentType':'application/json'}

```

Code 7.4: Route im Flask Webserver zur Beantwortung des Health Checks des ALB

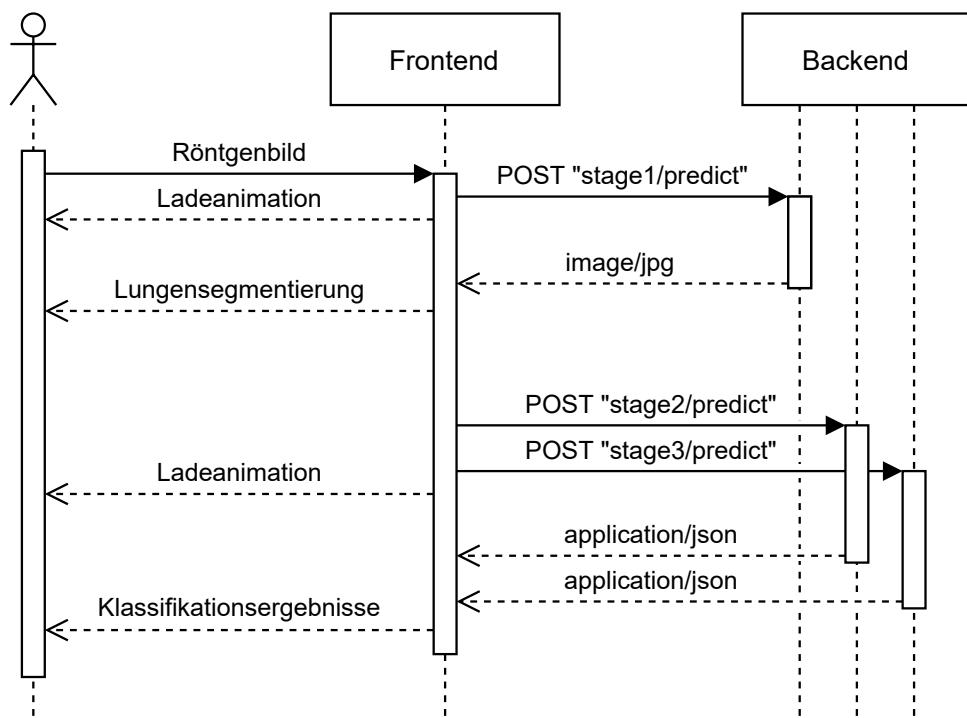


Abbildung 7.2: Sequenzdiagramm zur Beschreibung des typischen Ablaufs zwischen Nutzer, Webanwendung und Vorhersagemodellen bei der Analyse eines Röntgenbilds

Eine typische Kommunikation zwischen Nutzer, Webanwendung und Vorhersagemodellen ist im Sequenzdiagramm in Abbildung 7.2 zu sehen. Zu Beginn des Diagramms ist zu sehen, dass der Nutzer ein Röntgenbild in der grafischen Benutzeroberfläche hochlädt, woraufhin die Anwendung eine Ladeanimation anzeigt. Um weiterhin die Visualisierung von Zusatzinformationen in der grafischen Benutzeroberfläche zu demonstrieren, wird das Röntgenbild an das Modell zur Lungensegmentierung geschickt, welches unter der Route `stage1/predict` verfügbar ist. Die Anwendung war zu Beginn darauf ausgelegt, dass das Ergebnis der Lungensegmentierung für die weiteren Klassifikationen benötigt wird. Deshalb werden die weiteren Anfragen erst gesendet, wenn diese Operation abgeschlossen ist, auch wenn die aktuelle Version der Vorhersagemodelle die Segmentierung nicht mehr als Input benötigen. Da die Lungensegmentierung nur wenige Sekunden dauert, ist diese Konfiguration jedoch unproblematisch. Nachdem das Ergebnis eingetroffen ist, wird die erste Ladeanimation beendet und die Segmentierung dem Nutzer angezeigt.

Anschließend wird das Röntgenbild parallel an die beiden Vorhersagemodelle gesendet, da diese unabhängig voneinander agieren. Um dem Nutzer zu zeigen, dass sich das Bild in der Verarbeitung befindet, wird eine weitere Ladeanimation angezeigt. Die Ergebnisse beider Vorhersagen werden abgewartet, bevor eine Darstellung der Resultate in der Benutzeroberfläche geschieht. Das liegt daran, dass bei einer gesund klassifizierten Lunge das Ergebnis der Unterscheidung zwischen viraler und bakterieller Pneumonie nicht angezeigt werden soll. Da die Vorhersagen in etwa gleich lange benötigen und die Anfragen parallel versendet werden, stehen die Ergebnisse fast

zeitgleich zur Verfügung.

7.3 Datenschutz

Als Abschluss dieses Kapitels soll das Thema Datenschutz und Privatsphäre in Bezug auf das entwickelte System diskutiert werden. Bei der Nutzung von AWS Diensten kann der Nutzer auswählen, an welchem physischen Standort die Rechenleistung und Speicherkapazitäten zur Verfügung gestellt werden sollen. Die Standorte sind dabei in Regionen unterteilt, welche wiederum in sogenannte Availability Zones eingeteilt werden. Alle hier verwendeten Komponenten von AWS wurden für die Region eu-central-1 konfiguriert, was bedeutet, dass sie im Rechenzentrum in Frankfurt am Main ausgeführt werden. Somit ist sichergestellt, dass die Daten ausschließlich von Servern in Deutschland verarbeitet werden. Diese Server unterliegen den deutschen und europäischen Gesetzen und Richtlinien zum Datenschutz und Datensicherheit. Durch den 2018 in den USA in Kraft getretenen Cloud Act, müssen amerikanische Unternehmen den US-Behörden auch dann Zugriff auf gespeicherte Daten ermöglichen, wenn diese außerhalb der USA gespeichert sind. Laut Artikel 48 DSGVO kann dieses Gesetz jedoch nur angewendet werden, wenn ein Rechtshilfeabkommen zwischen einem EU-Staat und den USA besteht, was aktuell für keinen Mitgliedsstaat der Fall ist [18].

Die verschlüsselte Verbindung mittels CloudFront und dem Certificate Manager sorgt dafür, dass die sensiblen medizinischen Daten auf dem Übertragungsweg sicher sind. Da die Daten von den Vorhersagemodellen verarbeitet werden müssen, können dem Nutzer im Gegensatz zum Übertragungsweg keine Garantien für die ordnungsgemäße Behandlung der Daten gegeben werden. Das GitHub Repository des Projektes ist öffentlich zugänglich, was für eine gewisse Nachvollziehbarkeit und Transparenz der Implementierung sorgt. Trotz des verfügbaren Quellcodes ist es für den Nutzer jedoch nicht möglich, einen Einblick in die Vorgänge im ECS-Cluster und den Docker-Containern zu erhalten. Diese Situation aus Sicht des Datenschutzes die wohl größte Herausforderung für das hier beschriebene Assistenzsystem.

Für größere Einrichtungen wäre es denkbar, diesen Aufbau in der Cloud zugunsten des Datenschutzes auf ein lokales Rechenzentrum zu übertragen. Durch die Verwendung von weit verbreiteter Software wie Angular und der Portabilität der Docker-Images wäre dies ohne großen Aufwand umsetzbar. Besonders für kleinere Institutionen und Krankenhäuser sollte jedoch genauestens abgewogen werden, ob die Verarbeitung der Bilder ein Ausschlusskriterium für die Verwendung von Cloud-Technologien ist, da die hochperformante Infrastruktur von AWS viele Vorteile bietet. Die Bereitstellung von neuronalen Netzen innerhalb des Browsers ist zwar möglich, stellt für einen effizienten Einsatz in der realen Umgebung jedoch keine echte Alternative dar, was bereits in Abschnitt 6.4.2 ausführlich analysiert wurde.

Kapitel 8

Zusammenfassung und Ausblick

Im Folgenden werden die vorliegende Arbeit und ihre Ergebnisse kurz zusammengefasst. Weiterhin beschreibt dieses Kapitel sowohl allgemeine als auch für diese Arbeit spezifische, Limitationen des entwickelten Assistenzsystems. Abschließend wird ein Ausblick über die mögliche Weiterentwicklung und Optimierung des Systems gegeben.

8.1 Zusammenfassung der Arbeit

In dieser Arbeit wurde ein prototypisches Assistenzsystem zur Erkennung von Lungenentzündungen anhand von Thoraxröntgenbildern entwickelt. Die zugrundeliegenden Trainingsdaten wurden umfangreich analysiert und Missverhältnisse der Eigenschaften von Patienten mit und ohne Pneumonie dargelegt. Dieser Aspekt wurde bisher zu wenig thematisiert, obwohl Forschungsergebnisse deutlich machen, dass die Merkmale der Daten diskriminierende Vorhersagen neuronaler Netze provozieren können [75]. Im Rahmen der Arbeit konnten diese Missstände nicht nur aufgedeckt, sondern durch die Ausbalancierung der Klassen, Stratifikation der Eigenschaften und Kombination der Datensätze auch drastisch reduziert werden. Bei der Datenvorverarbeitung, Architektur und Entwicklung des Vorhersagemodells konnten aktuelle Erkenntnisse dieses noch relativ jungen Forschungsgebiets erfolgreich implementiert werden. Die Ergebnisse des Modells zeigen, dass durch die Verwendung kombinierter und stratifizierter Daten die Generalisierbarkeit der Vorhersagen verbessert wird.

Das umgesetzte CAD-System veranschaulicht das enorme Potential künstlicher Intelligenz in der Medizin. Die Benutzerfreundlichkeit der Anwendung hilft einer breiten Zielgruppe, auch ohne Vorkenntnisse der Medizin oder Informatik, einen Zugang zu diesem Thema zu finden. Die Vollständigkeit des Systems von der Dateneingabe in der grafischen Benutzeroberfläche, über das Konzept zur Bereitstellung, bis hin zur Darstellung der Vorhersageergebnisse, verdeutlicht wie die Technologie praxisnah eingesetzt werden könnte. Aufgrund der Systemarchitektur und der Implementierung der Webanwendung ist das entwickelte Assistenzsystem unabhängig von Hardware und Betriebssystem flexibel auf mobilen oder stationären Geräten einsetzbar.

8.2 Limitationen

Grundsätzlich sind die Ergebnisse dieser und weiterer Arbeiten sehr vielversprechend im Bezug auf die praktische Einsetzbarkeit von Vorhersagemodellen für Röntgenbilder. Da es sich der Anwendungsfall des Assistenzsystems im medizinischen Bereich befindet, gelten sehr hohe Qualitäts- und Sicherheitsstandards. Vor dem Einsatz eines solchen Systems sind somit diverse Zertifizierungen und medizinische Studien zu durchlaufen. Auch Ethische Aspekte sind beim Einsatz von KI eine zentrale Herausforderung. Daher ist ein Produktiveinsatz des Systems nicht ohne weiteres möglich.

Die Verfügbarkeit von Trainingsdaten ist ein weiterer limitierender Faktor von KI im Gesundheitssektor. Aufgrund von Datenschutzbestimmungen und der nötigen Infrastruktur zur Erzeugung der Datensätze ist eine breite Veröffentlichung von Trainingsdaten oft nicht möglich [58]. Selbst das Zusammentragen von etwa einer Millionen Röntgenbildern aus öffentlich verfügbaren Datensätzen in Kapitel 4 resultiert pro Krankheitslabel nur in einigen zehntausend Bildern. Zudem verwenden produktiv eingesetzte Systeme, wie das in Abschnitt 3.2 erwähnte qXR, häufig proprietäre Trainingsdaten, um einen Wettbewerbsvorteil zu erlangen [63].

Weiterhin ist die Erklärbarkeit der Vorhersagen aktuell nicht gegeben. Regionen, die besonders zur Klassifikation beigetragen haben könnten, beispielsweise anhand der Aktivierungsfunktionen ermittelt und durch eine Heatmap dargestellt werden. Somit könnten zumindest grobe Fehlklassifikationen durch die Präsenz von Artefakten oder medizinischen Hilfsgeräten in den Bildern ausgeschlossen werden.

8.3 Ausblick

Der in dieser Arbeit entstandene Quellcode ist öffentlich auf GitHub verfügbar [64]. Das System kann somit beliebig angepasst und weiterentwickelt werden.

Da sich diese Arbeit auf die Klassifikation von Pneumonien beschränkt, kann das Vorhersagemodell nur einen Teilaспект der Befundung von Thoraxröntgenbildern unterstützen (siehe Abschnitt 5.4.4). Eine Erweiterung des Systems für zusätzliche Krankheiten wäre daher besonders interessant. Die in Abschnitt 4.3.2 beschriebene Methode zur Stratifikation der Daten müsste dafür ebenfalls angepasst werden. Des Weiteren sollte der Zusammenschluss mehrerer Modelle zu einem sogenannten Ensemble in Betracht gezogen werden, um die Robustheit des Vorhersagemodells zu erhöhen.

Bei der Entwicklung des Assistenzsystems Chester verwendeten Cohen u. a. eine sogenannte out-of-distribution (OOD)-Detektion, um dem Anwender eine Rückmeldung darüber zu geben, ob das hochgeladene Bild für eine sinnvolle Vorhersage geeignet ist [13]. Dabei vergleichen sie das Eingabebild mit den Trainingsdaten und berechnen die Ähnlichkeit der Bilder. Somit können Fehleingaben wie natürliche Bilder oder Röntgenbilder anderer Körperregionen zurückgewiesen werden. Die Implementierung

einer OOD-Detektion wäre auch für das hier entwickelte System von Vorteil. Weiterhin könnte diese Methodik zur Datenvorverarbeitung genutzt werden. Bei der Durchführung des Vergleichs könnten beispielsweise minderwertige Bilder (siehe Abbildung 5.2c) und weitere Ausreißer aus den Datensätzen entfernt werden.

Die zuvor bereits erwähnte Nachvollziehbarkeit der Entscheidungen des Modells ist ein weiterer wichtiger Punkt für die Verbesserung des aktuellen Systems. Dabei bietet es sich besonders an, die Entscheidung des Modells auf visueller Ebene zu begründen. Für das Hervorheben von Regionen oder das Darstellen von Heatmaps ist die Webanwendung bereits vorbereitet. Der in Abschnitt 6.3.3 beschriebene Mechanismus zur halbtransparenten Darstellung der Lungensegmentierung über dem Röntgenbild könnte für diesen Zweck angepasst werden.

Literaturverzeichnis

- [1] Martin Abadi u. a. „TensorFlow: A System for Large-Scale Machine Learning“. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. USENIX Association, Nov. 2016, S. 265–283. ISBN: 978-1-931971-33-1.
- [2] Farah S. Ali u. a. „Diagnostic Radiology in Liberia: A Country Report“. In: *Journal of Global Radiology* 1.2 (Nov. 2015). doi: 10.7191/jgr.2015.1020.
- [3] Amazon Web Services. *Amazon Web Services - Server Hosting & Cloud Services*. 2021. URL: <https://aws.amazon.com/de/> (besucht am 03.05.2021).
- [4] Alan R. Aronson und François-Michel Lang. „An overview of MetaMap: historical perspective and recent advances“. In: *Journal of the American Medical Informatics Association* 17.3 (Mai 2010), S. 229–236. ISSN: 1067-5027. doi: 10.1136/jamia.2009.002733.
- [5] Erin P. Balogh, Bryan T. Miller und John R. Ball. *Improving Diagnosis in Health Care*. National Academies Press, 29. Dez. 2015. doi: 10.17226/21794.
- [6] Banco digital de Imagen Medica de la Comunidad Valenciana (BIMCV). *PadChest*. 2019. URL: <https://bimcv.cipf.es/bimcv-projects/padchest/> (besucht am 19.04.2021).
- [7] Bastian Benrath. „Cloudsparte AWS: Die Sonne hinter Amazons Wolken“. In: *FAZ.NET* (3. März 2019). ISSN: 0174-4909. URL: <https://www.faz.net/aktuell/wirtschaft/digitec/cloudsparte-aws-die-sonne-hinter-amazons-wolken-16067635.html> (besucht am 03.05.2021).
- [8] Blausen.com staff. „Medical gallery of Blausen Medical 2014“. In: *WikiJournal of Medicine* 1.2 (2014). doi: 10.15347/wjm/2014.010.
- [9] Jennifer Bryce u. a. „WHO estimates of the causes of death in children“. In: *The Lancet* 365.9465 (März 2005), S. 1147–1152. doi: 10.1016/s0140-6736(05)71877-8.
- [10] Aurelia Bustos u. a. „PadChest: A large chest x-ray image dataset with multi-label annotated reports“. In: *Med. Image Anal.*, 66 (2020), 101797 (22. Jan. 2019). doi: 10.1016/j.media.2020.101797. arXiv: 1901.07441v2 [eess.IV].
- [11] Yuvraj Chauhan. *NgxSpinner*. URL: <https://github.com/Napster2210/ngx-spinner> (besucht am 29.04.2021).
- [12] Cochran. *Sampling Techniques*. John Wiley & Sons, 18. Juli 1977. 448 S. ISBN: 047116240X.
- [13] Joseph Paul Cohen, Paul Bertin und Vincent Frappier. „Chester: A Web Delivered Locally Computed Chest X-Ray Disease Prediction System“. In: (31. Jan. 2019). arXiv: 1901.11210v3 [cs.CV].

- [14] Joseph Paul Cohen u. a. „On the limits of cross-domain generalization in automated X-ray prediction“. In: (6. Feb. 2020). arXiv: 2002.02497 [eess.IV].
- [15] Dina Demner-Fushman u. a. „Design and development of a multimodal biomedical information retrieval system“. In: *Journal of Computing Science and Engineering* 6.2 (2012), S. 168–177. URL: <https://doi.org/10.5626/JCSE.2012.6.2.168>.
- [16] Dina Demner-Fushman u. a. „Preparing a collection of radiology examinations for distribution and retrieval“. In: *Journal of the American Medical Informatics Association* 23.2 (Juli 2015), S. 304–310. ISSN: 1067-5027. doi: [10.1093/jamia/ocv080](https://doi.org/10.1093/jamia/ocv080).
- [17] Olaf Dössel. *Bildgebende Verfahren in der Medizin*. Springer Berlin Heidelberg, 2016. doi: [10.1007/978-3-642-54407-1](https://doi.org/10.1007/978-3-642-54407-1).
- [18] Europäisches Parlament und Rat der Europäischen Union. *Verordnung (EU) 2016/679*. 27. Apr. 2016. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32016R0679&qid=1619864153156> (besucht am 01.05.2021).
- [19] Carl E. Fabian. „Knowing right from left on X-rays: A way to minimize errors of laterality“. In: *Applied Radiology* 34.7 (2005), S. 19–24.
- [20] Google. *Angular*. Version 11.2.13. 2021. URL: <https://angular.io/> (besucht am 29.04.2021).
- [21] Google. *Angular Material UI component library*. Version 11.2.11. 2020. URL: <https://material.angular.io/> (besucht am 29.04.2021).
- [22] Google. *Material Design*. URL: <https://material.io/> (besucht am 29.04.2021).
- [23] Kaiming He u. a. „Deep Residual Learning for Image Recognition“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juni 2016, S. 770–778.
- [24] M. Hofer und N. Abanador. *The Chest X-ray: A Systematic Teaching Atlas*. Thieme Publishers Series. Thieme, 2007. ISBN: 9781588905543.
- [25] G. Höffken u. a. „Epidemiologie, Diagnostik, antimikrobielle Therapie und Management von erwachsenen Patienten mit ambulant erworbenen unteren Atemwegsinfektionen sowie ambulant erworbener Pneumonie – Update 2009“. In: *Pneumologie* 63.10 (Okt. 2009), e1–e68. doi: [10.1055/s-0029-1215037](https://doi.org/10.1055/s-0029-1215037).
- [26] A. S. Hollman und F. G. Adams. „The influence of the lordotic projection on the interpretation of the chest radiograph“. In: *Clinical radiology* 40.4 (1989), S. 360–364. doi: [10.1016/s0009-9260\(89\)80119-9](https://doi.org/10.1016/s0009-9260(89)80119-9).
- [27] Indiana Health Information Exchange, Inc. *Indiana Health Information Exchange*. URL: <https://www.ihiexchange.org/> (besucht am 15.04.2021).
- [28] Jeremy Irvin u. a. „CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison“. In: (21. Jan. 2019). arXiv: 1901.07031 [cs.CV].

- [29] *Information technology - Computer graphics and image processing - Portable Network Graphics (PNG): Functional specification*. Standard. International Organization for Standardization, 2004.
- [30] Stefan Jaeger u. a. „Two public chest X-ray datasets for computer-aided screening of pulmonary diseases“. In: *Quantitative Imaging in Medicine and Surgery* 4.6 (2014). ISSN: 2223-4306.
- [31] Alistair Johnson u. a. *MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0)*. 2019. doi: <https://doi.org/10.13026/8360-t248>.
- [32] Alistair E. W. Johnson u. a. „MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs“. In: (21. Jan. 2019). arXiv: 1901.07042v5 [cs.CV].
- [33] Alistair E. W. Johnson u. a. „MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports“. In: *Scientific Data* 6.1 (Dez. 2019). doi: 10.1038/s41597-019-0322-0.
- [34] Jerome P. Kassirer. „Our Stubborn Quest for Diagnostic Certainty“. In: *New England Journal of Medicine* 320.22 (1989), S. 1489–1491.
- [35] Alexander Ke u. a. „CheXtransfer: Performance and Parameter Efficiency of ImageNet Models for Chest X-Ray Interpretation“. In: *Proceedings of the Conference on Health, Inference, and Learning* (Apr. 2021). doi: 10.1145/3450439.3451867.
- [36] Daniel S. Kermany, Fang Zhang und Michael Goldbaum. *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*. 2018. doi: 10.17632/rscbjbr9sj.2. URL: <https://data.mendeley.com/datasets/rscbjbr9sj/2> (besucht am 09.04.2021).
- [37] Daniel S. Kermany u. a. „Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning“. English. In: *Cell* 172.5 (Feb. 2018), 1122–1131.e9. ISSN: 0092-8674, 1097-4172. doi: 10.1016/j.cell.2018.02.010. (Besucht am 11.04.2021).
- [38] H. M. Kulke. *Röntgendiagnostik von Thoraxerkrankungen*. De Gruyter, 2013. doi: doi:10.1515/9783110311501.
- [39] Narayanan Kumbakara. „Managed IT services: the role of IT standards“. In: *Information Management & Computer Security* 16.4 (Okt. 2008), S. 336–359. doi: 10.1108/09685220810908778.
- [40] Curtis P. Langlotz. *RadLex: A New Method for Indexing Online Educational Materials*. Nov. 2006. doi: <https://doi.org/10.1148/rg.266065168>.
- [41] Quoc Le und Tomas Mikolov. „Distributed Representations of Sentences and Documents“. In: *Proceedings of the 31st International Conference on Machine Learning*. Hrsg. von Eric P. Xing und Tony Jebara. Bd. 32. Proceedings of Machine Learning Research 2. Bejing, China: PMLR, Juni 2014, S. 1188–1196. url: <http://proceedings.mlr.press/v32/le14.html>.
- [42] Robert Leaman, Ritu Khare und Zhiyong Lu. „Challenges in clinical natural language processing for automated disorder normalization“. In: *Journal of Biomedical Informatics* 57 (Okt. 2015), S. 28–37. doi: 10.1016/j.jbi.2015.07.010.

- [43] Edward Loper und Steven Bird. „NLTK: The Natural Language Toolkit“. In: (17. Mai 2002).
- [44] Barbara Marston u. a. „Incidence of Community-Acquired Pneumonia Requiring Hospitalization“. In: *Archives of Internal Medicine* 157.15 (Aug. 1997), S. 1709–1718. doi: 10.1001/archinte.1997.00440360129015.
- [45] Medizino GmbH. 2021. URL: <https://medizino.de/medizintechnik/c> (besucht am 28.02.2021).
- [46] Medizino GmbH. *Neues Röntgen-Gerät kaufen - Ratgeber und Angebote*. 2021. URL: <https://medizino.de/medizintechnik/roentgen> (besucht am 28.02.2021).
- [47] Ministry of Health and Social Welfare. *Emergency Human Resources for Health Plan*. Government of Liberia, 2011.
- [48] Cristina Mitchell. *World Radiography Day: Two-Thirds of the World's Population has no Access to Diagnostic Imaging*. Pan American Health Organization / World Health Organization. 7. Nov. 2012. URL: https://www.paho.org/hq/index.php?option=com_content&view=article&id=7410:2012-dia-radiografia-dos-tercios-poblacion-mundial-no-tiene-acceso-diagnostico-imagen&Itemid=1926&lang=en (besucht am 25.02.2021).
- [49] Mozilla. *Web Storage API*. 30. Apr. 2021. URL: https://developer.mozilla.org/de/docs/Web/API/Web_Storage_API (besucht am 01.05.2021).
- [50] Atsushi Nambu. „Imaging of community-acquired pneumonia: Roles of imaging examinations, imaging diagnosis of specific pathogens and discrimination from noninfectious diseases“. In: *World Journal of Radiology* 6.10 (2014), S. 779. doi: 10.4329/wjrr.v6.i10.779.
- [51] National Institutes of Health Clinical Center. *CXR8. Chest X-ray Dataset of 14 Common Thorax Disease Categories*. 2017. URL: <https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345> (besucht am 12.04.2021).
- [52] National Library of Medicine. *Medical Subject Headings*. URL: <https://www.nlm.nih.gov/mesh/meshhome.html> (besucht am 15.04.2021).
- [53] National Library of Medicine. *Open-i Data Collection*. 9. Apr. 2021. URL: <https://openi.nlm.nih.gov/faq#collection> (besucht am 09.04.2021).
- [54] Nobel Media AB. *The Nobel Prize in Physics 1901*. 2021. URL: <https://www.nobelprize.org/prizes/physics/1901/summary/> (besucht am 20.02.2021).
- [55] Luke Oakden-Rayner. „Exploring Large-scale Public Medical Image Datasets“. In: *Academic Radiology* 27.1 (Jan. 2020), S. 106–112. doi: 10.1016/j.acra.2019.10.006.
- [56] Luke Oakden-Rayner. *Half a million x-rays! First impressions of the Stanford and MIT chest x-ray datasets*. 25. Feb. 2019. URL: <https://lukeoakdenrayner.wordpress.com/2019/02/25/half-a-million-x-rays-first-impressions-of-the-stanford-and-mit-chest-x-ray-datasets/> (besucht am 17.04.2021).

- [57] Open Source Vision Foundation. *OpenCV*. URL: <https://opencv.org/> (besucht am 30.05.2021).
- [58] Trishan Panch, Heather Mattie und Leo Anthony Celi. „The “inconvenient truth” about AI in healthcare“. In: *npj Digital Medicine* 2.77 (16. Aug. 2019). doi: 10.1038/s41746-019-0155-4.
- [59] Fabian Pedregosa u. a. „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830.
- [60] Yifan Peng u. a. „NegBio: a high-performance tool for negation and uncertainty detection in radiology reports“. In: *AMIA Summits on Translational Science Proceedings* 2018 (18. Mai 2018), S. 188–196. ISSN: 2153-4063.
- [61] Emma Pierson u. a. „An algorithmic approach to reducing unexplained pain disparities in underserved populations“. In: *Nature Medicine* 27.1 (13. Jan. 2021), S. 136–140. doi: 10.1038/s41591-020-01192-7.
- [62] Preetham Putha u. a. „Can Artificial Intelligence Reliably Report Chest X-Rays?: Radiologist Validation of an Algorithm trained on 2.3 Million X-Rays“. In: (19. Juli 2018). arXiv: 1807.07455 [cs.CV].
- [63] Qure.ai. *Automated Chest X-ray Interpretation – qXR*. URL: <https://qure.ai/qxr.html> (besucht am 28.05.2021).
- [64] Jan Raber, Daniel Friedmann und Sebastian Steindl. *chest_x_ray Repository*. 11. Mai 2021. URL: https://github.com/ai-explorers/chest_x_ray (besucht am 28.05.2021).
- [65] Radiological Society of North America. *Radlex Term Browser*. URL: <http://radlex.org/> (besucht am 15.04.2021).
- [66] Pranav Rajpurkar u. a. „CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning“. In: (14. Nov. 2017). arXiv: 1711.05225v3 [cs.CV].
- [67] Pranav Rajpurkar u. a. „MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs“. In: (11. Dez. 2017). arXiv: 1712.06957.
- [68] Recognizer Group GmbH. *Angular Gauge Chart*. URL: <https://github.com/recognizer/angular-gauge-chart> (besucht am 29.04.2021).
- [69] Frank B. Rogers. „Medical subject headings“. In: *Bulletin of the Medical Library Association* 51.1 (1963), S. 114–116.
- [70] Olaf Ronneberger, Philipp Fischer und Thomas Brox. „U-Net: Convolutional Networks for Biomedical Image Segmentation“. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, S. 234–241. ISBN: 978-3-319-24574-4. doi: https://doi.org/10.1007/978-3-319-24574-4_28.
- [71] Wilhelm Conrad Röntgen. „Über eine neue Art von Strahlen (Erste Mittheilung)“. In: *Sitzungsbericht der Physikalisch-Medizinischen Gesellschaft in Würzburg* (1895).

- [72] David Rosman u. a. „Imaging in the Land of 1000 Hills: Rwanda Radiology Country Report“. In: *Journal of Global Radiology* 1.1 (2015). doi: 10.7191/jgr.2015.1004.
- [73] Ami N. Rubinowitz, Mark D. Siegel und Irena Tocino. „Thoracic Imaging in the ICU“. In: *Critical Care Clinics* 23.3 (Juli 2007), S. 539–573. doi: 10.1016/j.ccc.2007.06.001.
- [74] Olga Russakovsky u. a. „ImageNet Large Scale Visual Recognition Challenge“. In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), S. 211–252. doi: 10.1007/s11263-015-0816-y.
- [75] Laleh Seyyed-Kalantari u. a. „CheXclusion: Fairness gaps in deep chest X-ray classifiers“. In: *Biocomputing 2021*. WORLD SCIENTIFIC, Nov. 2020, S. 232–243. doi: 10.1142/9789811232701_0022.
- [76] Stanford Machine Learning Group. *CheXpert: A Large Dataset of Chest X-Rays and Competition for Automated Chest X-Ray Interpretation*. 2019. URL: <https://stanfordmlgroup.github.io/competitions/chexpert/> (besucht am 12.04.2021).
- [77] Statistisches Bundesamt. *Todesursachenstatistik*. 2021. URL: <https://www-genesis.destatis.de/genesis/online?sequenz=tabelleErgebnis&selectionname=23211-0003> (besucht am 13.05.2021).
- [78] Christian Szegedy u. a. „Rethinking the Inception Architecture for Computer Vision“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juni 2016.
- [79] Technische Hochschule Mittelhessen, Institut für Medizinische Physik und Strahlenschutz (IMPS). *Konventionelles Röntgen*. URL: <https://www.thm.de/mni/forschung/institute-gruppen/imps/imps-strahlung-information/radiologie/modalitaeten/konventionelles-roentgen.html> (besucht am 22.02.2021).
- [80] The Pallets Projects und Armin Ronacher. *Flask*. URL: <https://palletsprojects.com/p/flask/> (besucht am 30.05.2021).
- [81] The pandas development team. *pandas-dev/pandas: Pandas 1.2.4*. 2021. doi: 10.5281/ZENODO.3509134.
- [82] The World Bank. *Liberia | Data*. 2021. URL: <https://data.worldbank.org/country/liberia> (besucht am 25.02.2021).
- [83] G. K. Wallace. „The JPEG still picture compression standard“. In: *IEEE Transactions on Consumer Electronics* 38.1 (1992), S. xviii–xxxiv. doi: 10.1109/30.125072.
- [84] Xiaosong Wang u. a. „ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases“. In: *IEEE CVPR 2017, pp. 2097-2106* (2017) (5. Mai 2017). doi: 10.1109/CVPR.2017.369. arXiv: 1705.02315v5 [cs.CV].
- [85] Karl Wiegers und Joy Beatty. *Software Requirements*. Microsoft Press, 15. Aug. 2013. ISBN: 0735679665.

- [86] World Health Organization. *Baseline country survey on medical devices*. URL: http://www.who.int/diagnostic_imaging/collaboration/ctpermill_14.jpg (besucht am 28.02.2021).
- [87] Lorna Young. *CT Departments Seek Workflow Improvements to Address Increased CT Utilization*. 3. Feb. 2020. URL: <https://imvinfo.com/ct-departments-seek-workflow-improvements-to-address-increased-ct-utilization/> (besucht am 28.02.2021).
- [88] Lorna Young. *Highlights from IMV 2019 X-ray/DR/CR Outlook Report*. 23. Apr. 2019. URL: <https://imvinfo.com/highlights-imv-2019-x-ray-dr-cr-outlook-report/> (besucht am 26.02.2021).
- [89] John R. Zech u. a. „Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study“. In: *PLOS Medicine* 15.11 (Nov. 2018). Hrsg. von Aziz Sheikh, e1002683. doi: 10.1371/journal.pmed.1002683.
- [90] Aston Zhang u. a. *Dive into Deep Learning*. 0.16.2. <https://d2l.ai>. 20. März 2021.

Abbildungsverzeichnis

1.1	Ursprüngliche Modellarchitektur zur Klassifikation von Pneumonien (Quelle des Röntgenbilds: [36])	2
2.1	Historische Aufnahme der Hand von Prof. Rudolf Albert von Koelliker aus 1896, aufgenommen im Physikalischen Institut der Universität Würzburg. (Quelle: [71])	6
2.2	Die Komponenten des Röntgensystems bei der Anfertigung einer Thoraxaufnahme (Quelle: [79])	7
2.3	Prinzipieller Aufbau von weit verbreiteten CT-Scannern (Quelle: [17, S. 140])	11
2.4	Länderspezifische Verbreitung von CT-Geräten, gemessen in Anzahl Geräte pro eine Millionen Einwohner (Quelle: [86])	12
3.1	Diagnoseprozess vom Auftreten eines gesundheitlichen Problems bis zur Behandlung, Konzipiert vom Komitee für diagnostische Fehler im Gesundheitswesen (Quelle: [5, S. 33])	14
3.2	Pneumonie im mittleren Bereich der rechten Seite des Thorax, erkennbar am homogen auftretenden Schatten. (Quelle: [38, S. 63])	18
3.3	Pneumonie im rechten oberen Bereich, erkennbar am runden Schatten. Die längliche Aufhellung innerhalb des Schattens deutet auf eine weitere Infektion hin, die unabhängig von der Pneumonie ist. (Quelle: [38, S. 72])	18
3.4	Vergleich der Röntgenbilder eines Patienten mit zentraler Pneumonie vor (3.4a) und 14 Tage nach (3.4b) der medikamentösen Behandlung. (Quelle: [38, S. 65f])	19
3.5	Anatomische Richtungen des Menschen in medizinischer Fachsprache (Quelle: [8])	21
3.6	Übersicht verbreiteter Perspektiven von Thorax Röntgenaufnahmen (Quelle: [10, S. 5])	24
3.7	Schematischer Vergleich der Größe des Herzens bei PA- und AP-Aufnahmen. Die Größe auf dem Röntgenbild (rotes Rechteck) hängt von der Distanz des Herzens (rote Ellipse) zur Strahlenquelle und zum Detektor ab. (Quelle: [10, S. 6])	25
4.1	Analyse und Filterung der Datensätze durch Jupyter Notebooks	28
4.2	Bilder von Patienten mit Pneumonie sind im CXR14 Datensatz deutlich häufiger in AP-Projektion aufgenommen als bei gesunden Patienten	32

4.3	Die Verteilung der medizinischen Hilfsgeräte im CheXpert Datensatz zeigt, dass das Label <i>Support Devices</i> bei gesunden Patienten häufiger vergeben wurde als bei Patienten mit Pneumonie	35
4.4	Übersicht der Altersverteilung von Patienten mit und ohne Pneumonie im CheXpert Datensatz in Form eines Histogramms	36
4.5	Verteilung der Projektionsart bei Samples mit und ohne Pneumonie im MIMIC-CXR Datensatz	39
4.6	Verteilung der medizinischen Hilfsgeräte bei Patienten mit und ohne Pneumonie im MIMIC-CXR Datensatz	40
4.7	Vergleich von zwei Röntgenbildern mit der Projektion COSTAL (Quelle: [6])	41
4.8	Verteilung des Alters der Patienten im PadChest Datensatz	42
4.9	Vergleich der Verteilung verschiedener Eigenschaften bei Samples mit und ohne Pneumonie im PadChest Datensatz	43
4.10	Die unterschiedliche Altersverteilung zwischen Patienten mit und ohne Pneumonie im PadChest Datensatz wird beim Ausbalancieren der Klassen übernommen, wenn keine Stratifikation angewendet wird	45
4.11	Altersverteilung der Patienten bei Einbeziehung der Altersgruppen im Stratifikationsprozess	49
4.12	Prozentuale Anteile der Datensätze im Gesamtdatensatz	50
5.1	Visualisierung des Datenflusses von LeNet, einem der ersten erfolgreich eingesetzten CNNs. Die Eingabe ist eine handgeschriebene Ziffer, die Zielvariable ein Vektor mit Wahrscheinlichkeiten für zehn Ziffern. (Quelle: [90, S. 250])	52
5.2	Die Diversität der Bilder sorgt für besondere Herausforderungen in der Datenvorverarbeitung (Quelle: [6])	54
5.3	Korrektur der Ränder und Seitenverhältnisse der Röntgenaufnahmen als Vorverarbeitungsschritte	55
5.4	AUROC-Metriken des Vorhersagemodells, welches auf dem Gesamtdatensatz trainiert wurde. Sowohl bei den 10 % interner Validierungsdaten als auch bei PadChest als externe Daten zeigt das Modell eine gute Generalisierbarkeit.	61
5.5	Konfusionsmatrizen für die Vorhersagen auf den internen und externen Validierungsdaten mit 1=Pneumonie und 0=Normal	62
5.6	ROC-Kurven und AUROC-Werte mehrerer Vorhersagemodelle, die auf unterschiedlichen Teilmengen des CheXpert Datensatz trainiert wurden. Die Ergebnisse mit stratifizierten Daten zeigen eine bessere Generalisierbarkeit bei der Vorhersage externer Daten.	63
6.1	Übersicht und Zusammenspiel der Komponenten der Webanwendung	71
6.2	Startseite der Webanwendung mit allgemeinen Informationen über das Projekt	75
6.3	Dialog zur Auswahl der Datei, die zur Analyse an das Backend gesendet werden soll	76

6.4	Ladeanimation zur Überbrückung der Wartezeit, bis die Serverantwort ankommt	76
6.5	Ergebnisse der Klassifikation der beiden hinterlegten Beispielbilder zur Demonstration der Röntgenbildanalyse	78
6.6	Patienten erscheinen in der Übersicht als Kachel. Über das Kontextmenü jeder Kachel können Aktionen ausgeführt werden.	79
6.7	Beispiele für das Erscheinungsbild der Webanwendung beim Verwenden eines mobilen Endgeräts	80
7.1	Architektschaubild der AWS Dienste zur Bereitstellung des Assistenzsystems	89
7.2	Sequenzdiagramm zur Beschreibung des typischen Ablaufs zwischen Nutzer, Webanwendung und Vorhersagemodellen bei der Analyse eines Röntgenbilds	91

Tabellenverzeichnis

3.1	Perspektivische Eigenschaften von Thoraxaufnahmen	23
4.1	Übersicht des Open-i Datensatz	29
4.2	Übersicht des CXR14 Datensatz	30
4.3	Übersicht des Pädiatrie Datensatz von Kermany u. a.	33
4.4	Übersicht des CheXpert Datensatz	34
4.5	Übersicht des MIMIC-CXR Datensatz	37
4.6	Übersicht des PadChest Datensatz	40
4.7	Übersicht der abhängig vom Datensatz berücksichtigten Eigenschaften bei der Stratifikation	48

Quellcodeverzeichnis

4.1	Methode zur Stratifikation der gesunden Patienten, basierend auf den Eigenschaften der Patienten mit Pneumonie	47
6.1	Beispielhafte Konfigurationsdatei der Webanwendung im JSON-Format, wobei sich die Zieladresse des Servers auf dem lokalen System befindet	72
6.2	Zusammenfassung der Methoden des <i>HttpService</i> zur Kommunikation mit dem Backend	73
6.3	Übersicht der verfügbaren Datenumformungen im <i>ConversionService</i> . .	74
7.1	Beispielhafte Struktur eines Endpunkts in Python mit einer Route zur Kommunikation mit dem neuronalen Netz	85
7.2	Dockerfile zur Erstellung eines Webserver Image	86
7.3	GitHub Action zur automatischen Bereitstellung neuer Versionen der Webanwendung	88
7.4	Route im Flask Webserver zur Beantwortung des Health Checks des ALB	90