

Aus der Klinik für Rheumatologie

der Heinrich-Heine-Universität Düsseldorf

Direktor: Univ.-Prof. Dr. med. Jörg Distler

Algorithmen basierte Analyse konventionell-radiografischer Krankheitszeichen entzündlicher Arthritiden unter Nutzung von Techniken der Künstlichen Intelligenz

Dissertation

zur Erlangung des Grades eines Doktors der Medizin

der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

Vorgelegt von
Matthias Benedikt Kors
2024

Widmung

Für meine Familie

Teile dieser Arbeit wurden veröffentlicht in:

Radke, K.L.; Kors, M.; Müller-Lutz, A.; Frenken, M.; Wilms, L.M.; Baraliakos, X.; Wittsack, H.-J.; Distler, J.H.W.; Abrar, D.B.; Antoch, G.; Sewerin, P.
Adaptive IoU Thresholding for Improving Small Object Detection: A Proof-of-Concept Study of Hand Erosions Classification of Patients with Rheumatic Arthritis on X-ray Images. Diagnostics 2023, 13, 104.

Zusammenfassung

Hintergrund: Erkrankungen des rheumatischen Formenkreises betreffen bis zu 1% der deutschen Bevölkerung und können unbehandelt zu starken Gelenkdeformationen und erheblicher Beeinträchtigung der Lebensqualität führen. Zur frühzeitigen Erkennung von strukturellen Schäden am Knochen wurden standardisierte Scoring Methoden entwickelt, welche eine hohe Reliabilität bei der Bewertung erosiver Veränderungen aufweisen. Jedoch ist die Anwendung dieser Methoden zeitaufwändig und bedarf speziell geschulten Personals, wodurch eine Anwendung außerhalb klinischer Studien erschwert wird. Eine Automatisierung dieser Methoden könnte die Qualität der Diagnostik erhöhen und Blindbereiche schließen.

Ziele: Das Ziel dieser Arbeit war die Entwicklung eines künstlichen neuronalen Netzwerks - convolutional neural network (CNN) zur reliablen und automatisierten Analyse und Bewertung erosiver Veränderungen in konventionell-radiografischen Bilddatensätzen.

Methoden: Aus den Datenbanken des Instituts für Diagnostische und Interventionelle Radiologie sowie der Klinik für Rheumatologie und des Hiller Forschungszentrums für Rheumatologie des Universitätsklinikums Düsseldorf wurden 300 Röntgenstandardaufnahmen beider Hände in anterior-posterior Projektion von 119 gesichert an einer rheumatoiden Arthritis (RA) bzw. Psoriasis Arthritis erkrankten Patient:innen gesammelt und mithilfe des validierten Sharp van der Heijde Scores (SvdH) auf Erosionsausprägung analysiert. Zur simultanen Erkennung und Analyse der Gelenkregionen wurde ein Netzwerk auf Basis des RetinaNet genutzt. Die Patient:innen wurden auf 3 Datensätze aufgeteilt: Trainings-, Validierungs- und Testdatensatz. Im Training wurde das Netzwerk optimiert und angepasst. Die manuell zugewiesenen SvdH-Scores und Gelenkkoordinaten dienten hierfür als Grundlage. Der Validierungsdatensatz diente der Überprüfung der Leistungsentwicklung des Netzwerks während des Trainings. Die Genauigkeit, den Gelenken die korrekten SvdH-Scores zuzuordnen, wurde schlussendlich am bisher ungesehenen Datensatz, dem Testdatensatz, überprüft. Im Rahmen der Auswertung wurde sowohl die Übereinstimmung der vom Netzwerk mit dem vom menschlichen Befunder zugeordneten Scores als auch die Übereinstimmung des menschlichen Befunders mit sich selbst nach 6 Monaten bestimmt.

Ergebnisse: Die Übereinstimmung der vom Netzwerk mit dem vom menschlichen Befunder zugeordneten Scores lag über alle Erosionsklassen bei 94%. Eine besonders hohe Übereinstimmung zeigte sich bei der Klassifikation von Gelenken mit einem SvH-Score von 0 bzw. 5. Im Vergleich der beiden Bewertungszeitpunkte lag die Übereinstimmungsgenauigkeit der manuellen Befundung bei 88,5%. Dabei lag die maximale Abweichung des menschlichen Befunders zwischen den zwei Bewertungszeitpunkten bei einem Erosionsgrad.

Diskussion: Wir konnten im Rahmen dieser Arbeit zeigen, dass unser Ansatz auf Basis des RetinaNet eine hohe Genauigkeit bei der Lokalisation und Bewertung erosiver Veränderungen aufweist. Insbesondere in anatomisch komplexen Regionen wie der Handwurzel konnte im Vergleich zu früheren Ansätzen eine hohe Genauigkeit erreicht werden. Eine Vergrößerung des Datensatzes könnte zur Erhöhung der Reliabilität des Systems im Verlauf beitragen. Auch die Implementierung der im SvdH Score ebenfalls bewerteten Gelenkspaltverschmälerung könnte darüber hinaus die Anwendbarkeit des Systems im klinischen Kontext erweitern.

Abstract

Background: Rheumatic diseases affect up to 1% of the German population and can lead to severe joint deformities and considerable impairment of quality of life if left untreated. For early detection of structural bone damage, standardized scoring methods have been developed which show high reliability in the assessment of erosive changes. However, the application of these methods is time-consuming and requires specially trained personnel, making them difficult to use outside clinical trials. Automation of these methods could increase the quality of diagnostics and close blind spots.

Objectives: The aim of this study was to develop a convolutional neural network (CNN) for reliable and automated analysis and evaluation of erosive change in conventionally radiographic image datasets.

Methods: From the databases of the Institute of Diagnostic and Interventional Radiology and the Department of Rheumatology and the Hiller Research Center for Rheumatology of the University Hospital Düsseldorf, 300 standard radiographs of both hands in anterior-posterior projection of 119 confirmed patients with rheumatoid arthritis (RA) or Psoriasis Arthritis (PsA) were collected and analyzed for erosive changes using the validated Sharp van der Heijde score (SvdH). A network based on RetinaNet was used for the simultaneous detection and analysis of the joint regions. The patients were divided into 3 data sets: Training-, Validation- and Test-Set. In Training, the network was optimized and adjusted. The manually assigned SvdH scores and joint coordinates served as the ground truth. The Validation-Set was used to check the performance evolution of the network during training. The accuracy of assigning the correct SvdH scores to the joints was finally verified on the previously unseen dataset, the Test-Set. As part of the evaluation, both the agreement of the scores assigned by the network with the scores assigned by the human assessor and the agreement of the human assessor with itself after 6 months were determined.

Results: The agreement of the scores assigned by the network with the scores assigned by the human diagnostician was 94% across all erosion classes. Particularly high agreement was seen in the classification of joints with SvH scores of 0 and 5, respectively. Comparing the two assessment time points, the agreement accuracy of the manual findings was 88.5%, while the maximum deviation of the human assessor between the two evaluation times was one degree of erosion.

Discussion: In this work we could show that our approach based on the RetinaNet has a high accuracy in localizing and evaluating erosive changes. Especially in anatomically complex regions like the carpus, a high accuracy could be achieved compared to previous approaches. Increasing the size of the data set could help to increase the reliability of the system even further. Additionally, the implementation of the joint space narrowing, which is also assessed in the SvdH score, could extend the applicability of the system in a clinical context.

Inhaltsverzeichnis

Zusammenfassung.....	I
Abstract.....	II
Abkürzungsverzeichnis.....	V
Abbildungsverzeichnis.....	VII
Tabellenverzeichnis.....	VII
1 Einleitung.....	1
1.1 Rheumatoide Arthritis.....	1
1.1.1 Definition, Klassifikation, Epidemiologie.....	1
1.1.2 Klinik.....	3
1.1.3 Diagnostik.....	4
1.1.4 Therapie.....	6
1.2 Psoriasis Arthritis.....	7
1.2.1 Definition, Klassifikation, Epidemiologie.....	7
1.2.2 Klinik.....	8
1.2.3 Diagnostik.....	9
1.2.4 Therapie.....	9
1.3 Maschinelles Lernen.....	11
1.3.1 Convolutional Neural Networks.....	11
1.3.2 Training.....	15
1.3.3 Validierung.....	16
1.3.4 Test.....	17
1.4 RetinaNet.....	18
1.5 Ziele der Arbeit.....	20
2 Material und Methoden.....	22
2.1 Datenakquise.....	22
2.2 Scoring nach Sharp van der Heijde.....	23
2.2.1 Erosionsscore.....	23
2.2.2 Gelenkspaltverschmälerungs-Score.....	25
2.3 Software und Hardware.....	27
2.4 Präprozessierung und Augmentation.....	29
2.5 Gelenklokalisierung und Erosionsbewertung durch das CNN.....	30
2.6 Bewertungskriterien.....	31

2.7	Statistische Analyse	31
2.8	Ethikvotum	32
3	Ergebnisse	33
3.1	Patientenmerkmale	33
3.2	Übereinstimmung von automatischer und manueller Gelenkbewertung	34
3.3	Zeitaufwand	35
4	Diskussion	36
4.1	Maschinelles Lernen in der Rheumatologie	36
4.2	Vorteile des RetinaNet	38
4.3	Besonderheiten und Limitationen	38
4.4	Ausblick	39
	Danksagung	40

Abkürzungsverzeichnis

a.p.	<i>anterior-posterior</i>
ACR	<i>American College of Rheumatology</i>
AMWF	<i>Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften</i>
bDMARD	<i>biologic Disease-modifying antirheumatic drug</i>
BSG	<i>Blutsenkungsgeschwindigkeit</i>
CASPAR	<i>Classification Criteria for Psoriatic Arthritis</i>
CCP	<i>Cyclisches Citrulliniertes Peptid</i>
CCP-AK	<i>CCP-Antikörper</i>
CDAI	<i>Clinical Disease Activity Index</i>
CMC	<i>Carpometacarpal</i>
CNN	<i>Convolutional Neural Network</i>
CRP	<i>C-reaktives Protein</i>
csDMARD	<i>conventional synthetic Disease-modifying antirheumatic drug</i>
CT	<i>Computertomographie</i>
DAS28	<i>Disease Activity Score 28</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DMARD	<i>disease-modifying anti-rheumatic drug</i>
EULAR	<i>European League Against Rheumatism</i>
FPN	<i>Feature Pyramid Network</i>
GEPARD	<i>German Psoriasis Arthritis Diagnostic Questionnaire</i>
GRAPPA	<i>Group for Research and Assessment of Psoriasis and Psoriatic Arthritis</i>
IP	<i>Interphalangeal</i>
JSN	<i>Joint Space Narrowing</i>
mAP	<i>mean average precision</i>
MCP	<i>Metacarpophalangeal</i>
MRT	<i>Magnetresonanztomographie</i>
NSAR	<i>Nicht Steroidale Antirheumatika</i>
PACS	<i>Picture Archiving and Communication System</i>
PIP	<i>proximales Interphalangeal</i>
PsA	<i>Psoriasis Arthritis</i>

PsoA	<i>Psoriasis Arthritis</i>
RA	<i>Rheumatoide Arthritis</i>
RC	<i>Radiocarpal</i>
RD	<i>Rheumatic Disease</i>
RF	<i>Rheumafaktor</i>
SC	<i>Scaphocapital</i>
SDAI	<i>Simple Disease Activity Index</i>
SvdH	<i>Sharp/van der Heijde</i>

Abbildungsverzeichnis

Abbildung 1: Beispiel für eine Convolution Operation	13
Abbildung 2: Beispiel für eine Pooling Operation	14
Abbildung 4: Beispielgelenke mit korrespondierendem Erosionsscore nach Sharp van der Hejde aus dem genutzten Datensatz.....	24
Abbildung 5: Bewertete Gelenke und Gelenkflächen im Rahmen des Sharp/van der Heijde Scores	26
Abbildung 6: Schematischer Aufbau des RetinaNet zur Lokalisation und Klassifizierung der Röntgenbilder der Hände	28
Abbildung 7: Repräsentative Darstellung der Augmentation	30
Abbildung 8: Ergebnisse der Erosionsbewertung in Form einer Konfusionsmatrix	35

Tabellenverzeichnis

Tabelle 1: 2010 American College of Rheumatology/European League Against Rheumatism Klassifikations Kriterien für rheumatoide Arthritis	2
Tabelle 2: CASPAR Kriterien.....	8
Tabelle 3: Erosions-Score nach Sharp van der Heijde.....	24
Tabelle 4: JSN-Score nach Sharp van der Heijde	25
Tabelle 5: Überblick der Patient:innen Charakteristika.....	33
Tabelle 6: Übersicht über die Verteilung der Erosionsgrade nach dem SvH-Erosionsscore auf der Grundlage der 768 Gelenke im Testdatensatz	34

1 Einleitung

1.1 Rheumatoide Arthritis

1.1.1 Definition, Klassifikation, Epidemiologie

Rheumatoide Arthritis (RA) bezeichnet eine chronisch-inflammatorische Erkrankung, welche sich insbesondere in Form von Gelenkschwellung und Schmerzen der Fingergrund- und Mittelgelenke manifestiert, wobei die Ätiologie noch nicht vollständig verstanden ist. Sowohl genetische als auch Umweltfaktoren scheinen eine wesentliche Rolle bei der Entstehung der rheumatoiden Arthritis zu spielen [2, 3].

Bei unzureichend oder zu spät behandelten Patient:innen kann die Erkrankung zu einer fortschreitenden Gelenkdestruktion mit starken Schmerzen, zunehmendem Funktionsverlust und konsekutiv progredienter Einschränkung der Lebensqualität sowie erhöhter Mortalität führen [4-6].

Die Prävalenz in der deutschen Bevölkerung wird je nach Falldefinition und Datengrundlage mit 0,55%-1,38% angegeben.

Mit einer Inzidenz von etwa 50–60/100.000 bei Frauen und bei 25–30/100.000 für Männer, zählt sie zu den häufigsten Erkrankungen des entzündlich-rheumatischen Formenkreises [7] und ist damit auch sozioökonomisch von hoher Bedeutung [8].

Die ursprünglichen Klassifikationskriterien für RA aus dem Jahr 1987 [9] wurden 2010 vom American College of Rheumatology (ACR) in Zusammenarbeit mit der European League Against Rheumatism (EULAR) überarbeitet und stützen eine Klassifikation auf vier gleichwertigen Kriterien: Schwellung bzw. Druckschmerz an den Gelenken, Serologie, Akute Phase Reaktion und Dauer der Symptome (Tabelle 1). Auch wenn die Klassifikationskriterien keine Diagnosekriterien darstellen und dafür weder entwickelt noch validiert wurden, werden diese, wie bei vielen anderen Erkrankungen auch, orientierend bei der Diagnose der Erkrankung angewendet.

Tabelle 1: 2010 American College of Rheumatology/European League Against Rheumatism Klassifikations Kriterien für rheumatoide Arthritis [8]

	Punkte
Ziel-Population: Patient:innen mit	
1) eindeutiger Synovialitis in mind. 1 Gelenk	
2) nicht durch andere Erkrankung erklärbare Synovialitis	
4 Kategorien (A-D); Addition der Punktwerte;	
Ein Wert von ≥ 6 ist nötig für die Diagnosestellung	
A. Schwellung durch Synovialitis bzw. Druckschmerz an Gelenken	
≤ 1 großes Gelenk	0
2-10 große Gelenke	1
1-3 kleine Gelenke	2
4-10 kleine Gelenke	3
>10 Gelenke (mind. 1 kleines Gelenk)	5
B. Serologie	
RF <u>und</u> CCP negativ	0
RF <u>oder</u> CCP niedrig positiv	2
RF <u>oder</u> CCP hoch positiv	3
C. Akute Phase Reaktion	
CRP <u>und</u> BSG normal	0
CRP <u>oder</u> BSG erhöht	1
D. Dauer	
< 6 Wochen	0
≥ 6 Wochen	1

- Kleine Gelenke: Handgelenk, Fingergrundgelenke (MCP), Fingermittelgelenke (PIP), Interphalangealgelenk des Daumens (IP), Zehengrundgelenke II - IV
- Große Gelenke: Schulter, Ellenbogen, Hüftgelenk, Knie, Sprunggelenk
- Niedrigpositiv: Erhöhter Titer, jedoch nicht über das Dreifache des laborabhängigen Normwertes
- Hochpositiv: Mindestens dreifach erhöhter Titer

1.1.2 Klinik

Die rheumatoide Arthritis kann sich durch eine Vielzahl unterschiedlicher Symptome präsentieren.

Patient:innen stellen sich jedoch häufig mit einer Reihe klassischer Leitsymptome vor, darunter Schmerzen, Schwellung und Steifheit, insbesondere in den Morgenstunden, im Bereich der Hand und Fingergelenke. Auch die kleinen Fußgelenke können im Rahmen der Erkrankung betroffen sein.

Bei etablierter Erkrankung ist ein polytopes und symmetrisches Befallsmuster der Gelenke typisch [10].

Darüber hinaus sind auch subfebrile Temperaturen sowie grippe-ähnliche Symptome regelmäßig Teil des Symptomkomplexes [11].

Die Abgrenzung zu arthrotischen Beschwerden kann vor allem bei Erstmanifestation der Erkrankung schwierig sein, wobei bei der Arthrose häufiger Finger Mittel- und Endgelenke (Typ Bouchard und Heberden) betroffen sind und deutlich seltener die RA typischen Metacarpophalangeal-Gelenke (MCP). Ruheschmerz und Morgensteifigkeit der symptomatischen Gelenkregion von über einer Stunde können in unklaren Fällen als richtungsweisende, klinische Kriterien herangezogen werden, um eine Abgrenzung zu erleichtern. [12]

In diesem Zusammenhang ist auch die Dauer der Arthralgien von entscheidender Bedeutung. Liegen die Symptome seit mindestens sechs Wochen vor, macht dies das Vorliegen einer frühen RA wahrscheinlicher, sofern diese nicht durch andere entzündlich-rheumatische Erkrankungen erklärbar sind [13].

1.1.3 Diagnostik

Die Diagnostik der RA stützt sich im Wesentlichen auf die Anamnese, körperliche Untersuchung, Bildgebung sowie Laboruntersuchungen. Dabei werden sowohl intra- als auch extraartikuläre Manifestationen berücksichtigt.

Die Beurteilung von Gelenkschwellung, Bewegungseinschränkung und systemischer Entzündungsaktivität sind wichtige Bestandteile des diagnostischen Prozesses [11].

Die Leitlinie der Deutschen Gesellschaft für Rheumatologie empfiehlt zur Erfassung der Krankheitsaktivität sowie zur Verlaufsbeobachtung eine regelmäßige Dokumentation von sogenannten Kompositscores, wie dem *Disease Activity Score 28 (DAS28)*, *Simple Disease Activity Index (SDAI)* oder *Clinical Disease Activity Index (CDAI)* [11]. Die Scores beinhalten eine Erfassung der Anzahl aller geschwollenen und druckschmerzhaften Gelenke sowie eine Einschätzung des Krankheitszustandes durch den Patienten und/oder durch den Untersucher. Darüber hinaus werden im SDAI und DAS28 laborchemische Entzündungsparameter wie das C-reaktive Protein (CRP) oder die Blutsenkungsgeschwindigkeit (BSG) in die Bewertung mit einbezogen.

In Zusammenschau mit weiteren Laboruntersuchungen und bildgebenden Verfahren wird so eine bessere Einschätzung der Krankheitsaktivität ermöglicht.

Im Rahmen der genannten bildgebenden Verfahren spielen vor allem das konventionelle Röntgen, der Gelenkulterschall sowie die Magnetresonanztomographie (MRT) eine entscheidende Rolle. CT-Untersuchungen sind seltener von Bedeutung, wobei auch hier zahlreiche wissenschaftliche Projekte die Wertigkeit dieser Untersuchungsmodalität untersuchen.

Die MRT zeichnet sich durch eine hohe Auflösung und Sensitivität in Bezug auf Synovialitis, Tendovaginitis, Knochenmarködeme und Erosionen aus [14].

Das konventionelle Röntgen weist zwar in Bezug auf frühe knöchernen Veränderungen eine geringere Sensitivität auf als die MRT [15, 16], ist aber aufgrund breiter Verfügbarkeit, hoher Reliabilität und niedriger Kosten weiterhin ein zentrales Element der Diagnostik und Verlaufsbeurteilung der rheumatoiden Arthritis [17]. Die hohe örtliche Auflösung des konventionellen Röntgens ermöglicht zudem eine gute objektive Visualisierung des Knochens und der in Folge der Erkrankung auftretenden, knöchernen Destruktion. Unter Zuhilfenahme

verschiedener Scoringmethoden, wie dem Sharp van der Heide Score, können darauf aufbauend die Ergebnisse des Röntgens quantifiziert und verlaufskontrolliert werden.

1.1.4 Therapie

Das Therapiekonzept der RA basiert sowohl auf medikamentösen als auch auf nicht-medikamentösen Interventionen mit dem Ziel der Krankheitsremission [11].

Eine zentrale Säule der Therapie stellen krankheitsmodifizierende, antirheumatische Medikamente (engl.: Disease-Modifying Anti-Rheumatic Drugs, kurz: DMARDs) dar. Sie sind von entscheidendem therapeutischem Wert, da sie über die Wirkung auf die Symptome der RA hinaus die Gelenkzerstörung deutlich verlangsamen und in vielen Fällen heutzutage sogar stoppen können [18].

Der frühestmögliche Einsatz einer wirksamen, krankheitsmodifizierenden, antirheumatischen Therapie kann sowohl die Progression der RA positiv beeinflussen als auch die Chancen auf Remission erhöhen [19, 20].

Darüber hinaus werden Glucocorticoide und Analgetika unterschiedlicher Klassen (NSAR, Opioid-/Nicht-Opioid Analgetika) zur Symptom- und Entzündungsreduktion eingesetzt, wobei gerade die Rolle der Glucocorticoide auf Grund der vielen ungünstigen Langzeitnebenwirkungen (z.B. Osteoporose, Hypertonie, Infektionen) kritisch hinterfragt und deren Einsatz nur über kurze Zeit hinweg empfohlen wird [11].

Nicht-medikamentöse Maßnahmen umfassen u.a. Physio-, Ergo-, Bewegungs- und Sporttherapie sowie Physikalische Therapie. Auch psychologische Interventionen haben sich als wirksame Therapiekomponente erwiesen [21].

1.2 Psoriasis Arthritis

1.2.1 Definition, Klassifikation, Epidemiologie

Die Psoriasis Arthritis (PsA) ist eine chronisch-entzündliche Autoimmunerkrankung in Form einer Rheumafaktor negativen, meist asymmetrischen Oligo- oder Polyarthritis der kleinen und großen Gelenke sowie der unteren Extremitäten.

Neuesten Analysen deutscher Gesundheitsdaten zufolge liegt die Gesamtprävalenz der PsA bei ca. 0,1-0,2% der deutschen Bevölkerung [22].

Etwa 10-20% der Patient:innen mit einer Psoriasis vulgaris entwickeln im Laufe der Erkrankung eine Gelenkmanifestation im Sinne einer PsA [22, 23]. Sie stellt damit die wichtigste extrakutane Manifestation der Psoriasis dar.

In circa 68% der Fälle tritt die Arthritis dabei über 10 Jahre nach Beginn der kutanen Symptomatik auf [24].

Ein Teil der Patient:innen zeigt jedoch schon vor der Entwicklung kutaner psoriatischer Symptome erste Gelenkbeschwerden [24].

Kommt es zu einer Gelenkbeteiligung vor ersten kutanen Manifestationen, ist die PsA gelegentlich schwer von anderen entzündlichen Gelenkerkrankungen zu unterscheiden. Aus diesem Grund wurden verschiedene Klassifikationen entwickelt, welche eine bessere Abgrenzung ermöglichen [24-28].

Insbesondere die „*Classification criteria for psoriatic arthritis*“ (CASPAR) haben sich als einfach anzuwenden, vergleichsweise spezifisch und sensitiv erwiesen [29].

Tabelle 2: CASPAR Kriterien [29]

	<u>Punkte</u>
Eine PsA liegt vor bei entzündlicher Gelenkerkrankung (Gelenk, Wirbelsäule oder Haut) mit ≥ 3 Punkten aus den folgenden 5 Kategorien:	
1. Nachweis einer aktuellen Psoriasis	2
• Psoriasis in der Anamnese (falls keine aktuelle Psoriasis vorliegt)	1
• Psoriasis in der Familienanamnese (falls keine aktuelle Psoriasis und keine Psoriasis in der Anamnese vorliegen)	1
2. Typische psoriatische Nageldystrophie, einschließlich Onycholyse, Lochfraß und Hyperkeratose, in der aktuellen körperlichen Untersuchung	1
3. Negativer Rheumafaktor	1
4. Aktuelle Daktylitis	1
• Daktylitis in der Anamnese (falls keine aktuelle Daktylitis vorliegt)	1
5. Radiologischer Nachweis juxtaartikulärer Knochenneubildung, die als schlecht definierte Verknöcherung in der Nähe der Gelenkränder (aber ohne Osteophytenbildung) auf einfachen Röntgenbildern der Hand oder des Fußes erscheint	1

1.2.2 Klinik

Die Symptomatik der PsA ist äußerst variabel. Periphere Arthritis, Spondylitis, Enthesitis und Daktylitis sind im Rahmen der Erkrankung besonders häufig. Prädilektionsstellen der Gelenksymptomatik sind das Kniegelenk, die Sprunggelenke sowie die proximalen Interphalangealgelenke (PIP) und distalen Interphalangealgelenke (DIP) der Hände und Füße. Dabei sind sowohl mono-, oligo- und polyarthritische Formen als auch eine Wirbelsäulenmanifestation möglich [30]. Eine verminderte körperliche Leistungsfähigkeit sowie Beeinträchtigung der psychischen Gesundheit und konsekutiv eingeschränkte Lebensqualität sind häufig die Folge [31].

1.2.3 Diagnostik

Analog zur Diagnostik der RA werden bei der PsA neben der Anamnese und körperlichen Untersuchung auch bildgebende Verfahren und Laboruntersuchungen zur Objektivierung der Entzündungsaktivität herangezogen.

Die Leitlinie der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AMWF) empfiehlt neben der Bewertung klinischer Entzündungszeichen wie Rötung, Schwellung, Überwärmung, Schmerzen und Funktionsstörung bei klinisch vorliegender Arthritis bzw. Daktylitis einen Nachweis der Gelenksveränderungen im Gelenkulterschall, Röntgen oder MRT [32].

Des Weiteren wurden aufgrund einer hohen Prävalenz subklinischer PsA Verläufe Patientenfragebögen als Screening Instrument entwickelt. Darunter hat sich insbesondere der „*German Psoriasis Arthritis Diagnostic Questionnaire*“ (GEPARD) - ein validierter Screeningfragebogen in deutscher Sprache - als einfache und sensitive Methode zur Identifikation von Patient:innen mit PsA herausgestellt [33, 34].

1.2.4 Therapie

Zur Behandlungsplanung der PsA können sowohl die Leitlinie der EULAR (European League Against Rheumatism) als auch der GRAPPA (Group for Research and Assessment of Psoriasis and Psoriatic Arthritis) herangezogen werden [35, 36]. In einer Gegenüberstellung von Gossec et al. wurden die Gemeinsamkeiten und Unterschiede der Therapieempfehlungen herausgearbeitet [37]. Die Empfehlungen sind überaus umfangreich und in diesem Zusammenhang sollen nur die wesentlichen Aspekte dargestellt werden.

Sowohl EULAR als auch GRAPPA empfehlen eine interdisziplinäre Versorgung der Patient:innen und eine gemeinsame Behandlungsplanung. Die EULAR legt den Fokus mehr auf die rheumatologischen Aspekte der Erkrankung, während die GRAPPA interdisziplinär auch die dermatologischen Aspekte der PsA stärker priorisiert. Die übergreifenden Prinzipien der beiden Publikationen sind jedoch ähnlich. So empfehlen beide Gesellschaften eine *Treat-to-Target* Strategie mit sukzessiver Anpassung der Therapie [37]. Der frühe Einsatz von bDMARDs, wie z.B. TNF-, IL17- oder IL23-Inhibitoren bei unzureichendem Ansprechen auf eine csDMARD Therapie, ist dabei ein zentrales Prinzip beider Empfehlungen [38, 39].

Darüber hinaus sind auch nicht-medikamentöse Maßnahmen wie Physio-, Ergo-, Bewegungs- und Sporttherapie sowie Physikalische Therapie wichtige Bestandteile der Therapie der PsA. Analog zur Therapie der RA dienen sie vor allem der Symptomkontrolle der Gelenkmanifestationen [30].

Als primäres Ziel der therapeutischen Ansätze definiert die EULAR eine Krankheitsremission bzw. möglichst niedrige Krankheitsaktivität [38].

1.3 Maschinelles Lernen

Maschinelles Lernen (engl. machine learning) bezeichnet ein Teilgebiet der Informatik, welches sich mit der Entwicklung und Optimierung von statistischen Modellen zur Analyse komplexer Datensätze beschäftigt [40]. Aufgrund der vielfältigen Anwendungsmöglichkeiten dieser Technologie hat das Forschungsfeld des maschinellen Lernens in den vergangenen Jahren einen großen Aufschwung erlebt [41].

Die Anwendungsschwerpunkte liegen im wissenschaftlichen Kontext vor allem in der Erkennung und Analyse komplexer Strukturen und Zusammenhänge in mehrdimensionalen Datensätzen. Darunter fallen z.B. die Vorhersage von Molekülinteraktionen in der pharmakologischen Forschung [42] oder die Rekonstruktion neuronaler Strukturen und Interaktionen [43].

Einen wesentlichen Schwerpunkt der aktuellen Forschung und Anwendung in der Medizin stellen sogenannte *Deep Learning* Algorithmen zur Erkennung und Analyse medizinischer Bilddatensätze wie CT-, MRT oder Röntgenbildern dar [44-51].

Unter *Deep learning* werden Modelle mit mehreren Verarbeitungsschichten verstanden, welche auf diese Weise zur Verarbeitung großer und komplexer Datensätze in der Lage sind und sich daher insbesondere für Anwendungen im Bereich der Objekterkennung eignen.

Convolutional Neural Networks (CNN) sind dabei eine von vielen unterschiedlichen Formen von *Deep Learning* Algorithmen und haben sich als dominante Methode im Bereich der *Computer Vision* etabliert.

1.3.1 Convolutional Neural Networks

Ein CNN bezeichnet einen *Deep Learning* Algorithmus, der auf die Verarbeitung von gridartigen Datensätzen in Form einer Matrix (z.B. Pixel oder Voxel) ausgelegt ist [52].

Die Architektur ist hierbei an die Organisation des visuellen Cortex von Säugetieren angelehnt [53, 54] und kann grob in 3 Schichten unterteilt werden: *Convolutional Layer*, *Pooling Layer* und *Fully-connected Layer* [55, 56].

Die Faltungsschichten (engl. Convolutional Layer) dienen der Merkmalsextraktion aus der Eingabe, wie z.B. Knochenkonturen oder Gelenkflächen. Analog zum biologischen Vorbild des rezeptiven Feldes werden abschnittsweise die visuellen Informationen aus der Eingabe extrahiert.

Hierfür wird eine Filtermatrix (engl. Kernel) auf eine Eingabe, auch *Input Tensor* genannt, angewendet (siehe Abb. 1) [57]. Bei der Eingabe handelt es sich im medizinischen Kontext meist um zwei oder dreidimensionale Bilddatensätze wie Röntgen, CT oder MRT-Bilder. Im Falle eines Röntgenbildes entspricht der Zahlenwert der Pixel einer spezifischen Graustufe.

Die Filtermatrix wird auf einen in der Größe korrespondierenden Ausschnitt der Eingabe angewandt und das innere Produkt der beiden gebildet. Die Summe der Produkte aus Bildausschnitt und Filtermatrix bildet die Ausgabe (engl. Output) der Faltungsoperation. Die Ausgabe wird in der Merkmalskarte (engl. Feature Map) festgehalten [58]. Je nach Konfiguration des CNN's wandert die Filtermatrix anschließend eine vorgegebene Anzahl an Pixeln weiter über das Bild und wiederholt diesen Vorgang bis die gesamte Eingabe erfasst wurde [56]. Die ausgegebene Merkmalskarte enthält schlussendlich die Bildinformationen in komprimierter und abstrahierter Form.

Dieser Vorgang kann je nach Art und Konfiguration des CNN's unterschiedlich oft wiederholt werden. Mit der Anzahl der Wiederholungen steigt die Größe des rezeptiven Feldes und die Komplexität der extrahierten Merkmale konsekutiv an [59].

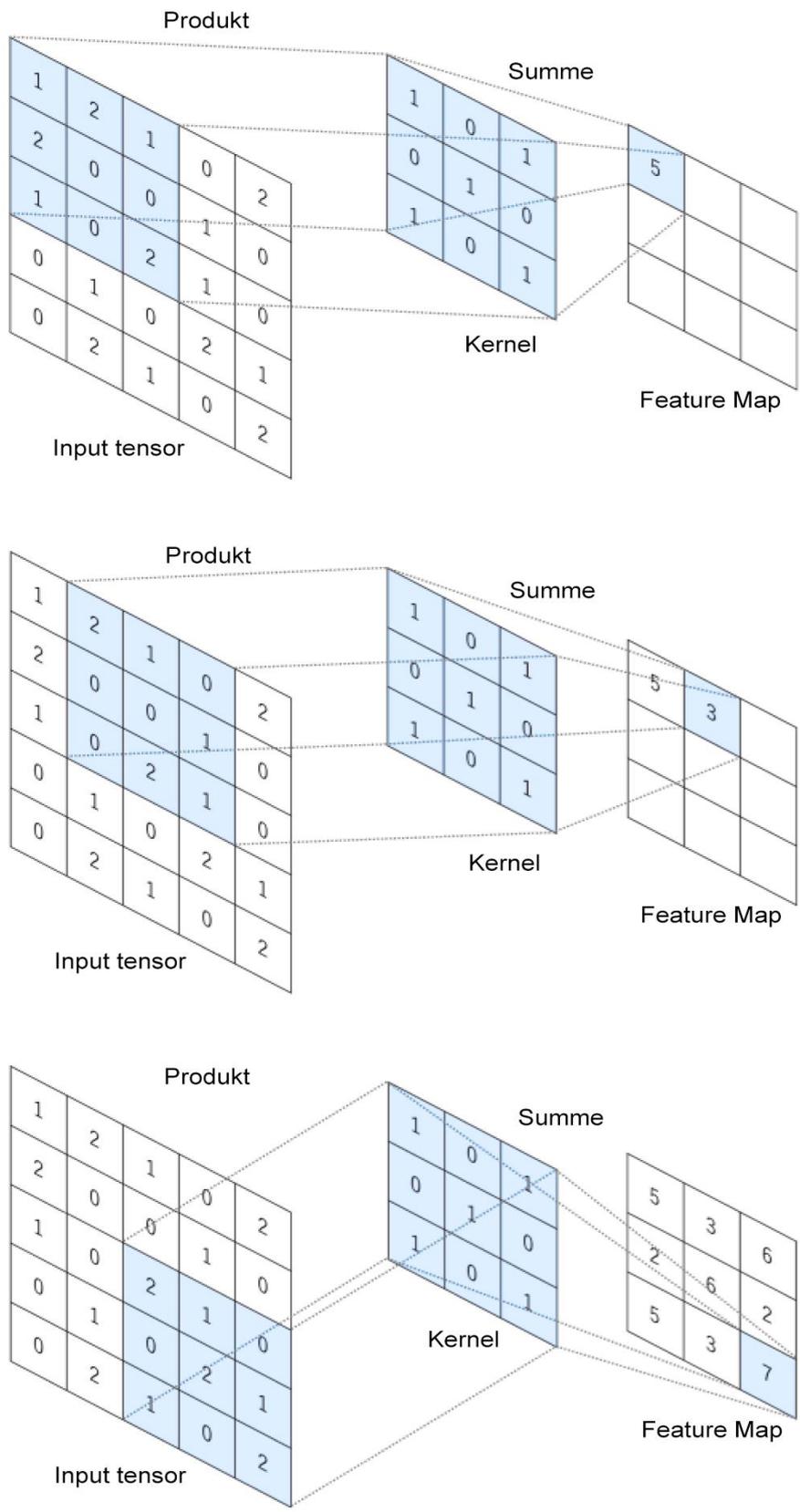


Abbildung 1: Beispiel für eine Convolution Operation

(modifiziert nach Yamashita et al., 2018)

Eine Filtermatrix wird auf die Eingabe angewendet und das Produkt über jedem Punkt errechnet. Die Summe ergibt schlussendlich die Ausgabe, in diesem Fall die *Feature Map*.

Das nächste Element bilden die Zusammenlegungsschichten (engl. Pooling Layer). Dabei wird die Größe der Eingabematrix reduziert, welches als Heruntertaktung (engl. Downsampling) bezeichnet wird. Eine Form der *Pooling Operation* ist das *Max Pooling*. Im Rahmen des *Max Pooling* wird nur der höchste Wert des betrachteten Bereiches für die weiteren Berechnungsschritte übernommen (siehe Abb. 2).

Folglich werden die wichtigsten Merkmale eines Bereiches hervorgehoben und Bildrauschen reduziert [60].

Dies hat zwar insgesamt einen Informationsverlust zur Folge, erhöht jedoch die Verarbeitungsgeschwindigkeit und vermindert den Hardwarebedarf. Es ermöglicht zudem tiefere Netzwerke zu erzeugen, die zur Verarbeitung komplexerer Datensätze fähig sind.

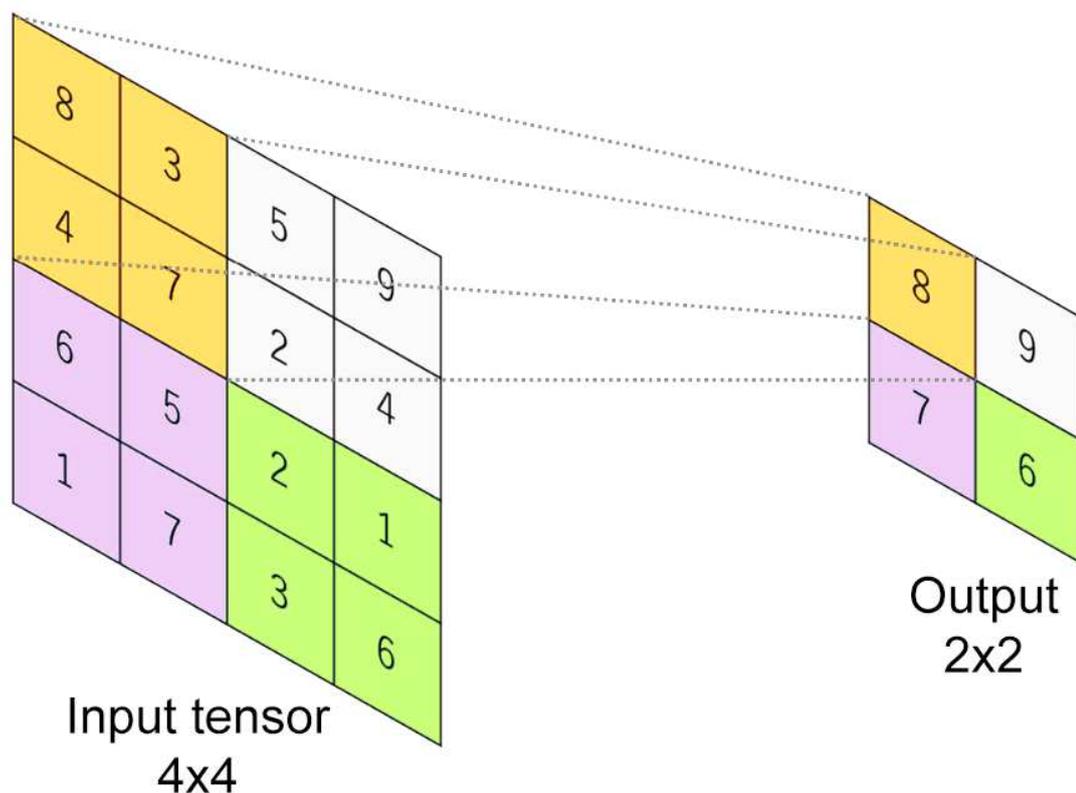


Abbildung 2: Beispiel für *Pooling Operation*
(modifiziert nach Yamashita et al., 2018)

Den abschließenden Teil des CNN bilden die vollvernetzten Schichten (engl. Fully-connected Layer). Diese Schicht ist nur dann ein Teil der Netzwerkarchitektur, wenn eine Klassifikation der gesammelten Objektinformationen erfolgen soll. Die in den *Convolutional* und *Pooling Layern* erzeugten mehrdimensionalen Merkmalskarten werden dafür in einen eindimensionalen Vektor umgewandelt und in die vollvernetzten Schichten eingespeist. Dabei gehen die Positionsdaten verloren und übrig bleiben positionsunabhängige Objektinformationen [61].

Je nach Aufgabe des CNN findet in diesem Abschnitt die Zuordnung der analysierten Eingabe zu den im Vorhinein definierten Kategorien bzw. Objektklassen statt. Im Rahmen dieser Arbeit waren dies die Zuordnung zur Gelenkregionen der Hände sowie die Ausprägung der Erosionen.

1.3.2 Training

Der erste Schritt in der Entwicklung eines CNN's ist das Training. Es lassen sich grob drei Ansätze differenzieren, welche sich in ihrer Herangehensweise an das Training eines Netzwerks unterscheiden:

Überwachtes Lernen, Unüberwachtes Lernen und bestärkendes Lernen [40].

Das überwachte Lernen bezeichnet einen Prozess des maschinellen Lernens, bei dem die Zielausgabe, im Falle dieser Arbeit der Grad der Erosion und die Gelenkposition bereits bekannt ist und dem Netzwerk als Grundwahrheit zur Verfügung gestellt wird. Genauer gesagt kommen Trainingsdatensätze zum Einsatz, welche im Vorhinein mit Informationen zu den enthaltenen Objekten und ihrer entsprechenden Position versehen wurden [62, 63].

Soll also ein Röntgenbild der Hand auf Erosionsausprägung analysiert werden, enthält der Trainingsdatensatz sowohl die Angabe über die Lokalisation der Gelenke als auch ihren Erosionsgrad.

Ziel des überwachten Lernens ist es, anhand der eingegebenen Informationen ein Modell zu erstellen, welches die gewünschte Ausgabe erzeugt.

Dafür werden die Netzwerkparameter im Verlauf des Trainings über mehrere Epochen angepasst und die daraus resultierenden Ergebnisse mit der Grundwahrheit abgeglichen [64]. Eine Epoche ist dabei definiert als ein vollständiger Durchlauf des gesamten Trainingsdatensatzes [65].

Beim unüberwachten Lernen stehen dem Netzwerk hingegen nur unbeschriftete Daten (engl: unlabeled Data) zur Verfügung [59]. Es steht somit keine gewünschte

Ausgabe zur Verfügung, anhand derer die Netzwerkparameter angepasst werden können. Die Eingabe wird daher auf inhärente Strukturen und Muster untersucht und abhängig davon in Gruppen (engl.: Cluster) eingeordnet [63].

Auf diesem Weg können z.B. große Datensätze geordnet und somit besser weiterverarbeitet werden.

Ein weiterer Ansatz des Netzwerktrainings ist die Methode des bestärkenden Lernens, welche auf dem Prinzip der Belohnungsmaximierung basiert [66]. Der Methodik liegen drei wesentliche Komponenten zugrunde: Der Agent, die Strategie und die Umgebung. Im Rahmen eines Markov Entscheidungsprozesses entwickelt der Agent, in diesem Fall der Algorithmus, eine Strategie, welche das maximale Belohnungssignal aus der Interaktion mit der Lernumgebung generiert [66, 67]. Der Wert einer Aktion wird dabei durch eine sogenannte Belohnungsfunktion (engl. reward function) festgelegt [68].

Auf diese Weise können komplexe, mehrschrittige Entscheidungsprozesse automatisiert und optimiert werden [69].

1.3.3 Validierung

Der Validierungsdatensatz (engl. Validationset) nimmt eine wichtige Stellung in der Entwicklung eines CNN's ein. Er dient der Überwachung der Leistung des Netzwerks während des Trainingsprozesses [40].

Im Rahmen des Trainingsprozesses kommt es oft zu einem Phänomen, welches als Überanpassung (engl. Overfitting) bezeichnet wird. Eine Überanpassung tritt insbesondere bei sehr langer Trainingszeit und bei Datensätzen hoher Komplexität auf [70].

Die Netzwerkparameter werden in diesem Fall so spezifisch auf den Trainingsdatensatz angepasst, dass eine hohe Übereinstimmung der Eingabe und erwünschten Ausgabe erreicht wird. Diese basiert jedoch weniger auf der Identifikation generalisierbarer Merkmale, sondern auf hoch spezifischen, nicht auf andere Datensätze übertragbaren Eigenschaften des Datensatzes [71]. Überprüft man folglich die Netzwerkleistung auf noch ungesehenen Daten, so lassen sich im Vergleich enorme Leistungseinbußen feststellen [72].

Werden hingegen hohe Genauigkeiten sowohl auf dem Trainings- als auch dem Validierungsdatensatz erreicht ist eine Überanpassung unwahrscheinlicher und ein

Netzwerkmodell, welches für die Anwendung auf unbekanntem Datensätzen geeignet ist, kann gut identifiziert werden.

1.3.4 Test

Der Testdatensatz bildet den letzten Schritt in der Erstellung eines neuronalen Netzwerkes.

Wurde auf dem Trainingsdatensatz erfolgreich ein Netzwerkmodell erstellt, so muss im Anschluss die Übertragbarkeit auf weitere, noch unbekanntem Datensätze überprüft werden. Aus diesem Grund ähnelt der Testdatensatz dem Trainingsdatensatz hinsichtlich der Struktur und der Verteilungswahrscheinlichkeiten, beinhaltet jedoch ausschließlich zuvor noch nicht gesehene Daten [40].

Welches Netzwerkmodell für die Testung ausgewählt wird, hängt von der entsprechenden Leistung im Validierungsdatensatz ab.

Werden schließlich auf den Testdaten ähnlich hohe Genauigkeiten wie auf den Trainingsdaten erreicht, deutet dies auf eine gute Generalisierung des Modells hin. Folglich ist das Modell auch für eine Anwendung außerhalb der bekannten Datensätze geeignet.

Die Übereinstimmung von Grundwahrheit und Vorhersage des Netzwerkmodells kann im Anschluss z.B. durch eine Konfusionsmatrix oder Wahrheitsmatrix visualisiert werden (siehe Abb. 8).

1.4 RetinaNet

Eine große Herausforderung im Bereich des maschinellen Lernens und insbesondere in der Entwicklung von CNN's ist das sogenannte Klassenungleichgewicht (engl. class imbalance).

Die Regionen des Bildes, welche nützliche Informationen für die Analyse beinhalten, stellen meist nur einen Bruchteil der Gesamtfläche des Bildes dar. So machen die Gelenkregionen in konventionellen Röntgenaufnahmen der Hände nur einen geringen Teil der gesamten Bildfläche aus. Folglich sind relevante Bereiche im Verhältnis zum Hintergrund stark unterrepräsentiert. Im Rahmen des Trainingsprozesses kann die große Anzahl an nicht relevanten Bildabschnitten die Detektionsleistung eines Netzwerks enorm verschlechtern [73].

In den vergangenen Jahren wurden aus diesem Grund verschiedene Ansätze entwickelt, um den Einfluss dieses Effektes zu reduzieren.

Ein besonders vielversprechender Ansatz ist das RetinaNet. Dieses wurde 2017 von Tsung-Yi Lin et al. im Rahmen ihrer Forschung in der *Facebook AI Research Group* vorgestellt [74, 75]. Die Netzwerk Architektur zeichnet eine hohe Leistung bei der Erkennung von kleinen Objekten und dichten Bildbereichen aus [76].

Die vorgestellte Variante des RetinaNet setzt sich aus 4 Hauptbestandteilen zusammen: Rückgrat (engl. Backbone), Merkmalspyramiden Netzwerk (engl. Feature Pyramid Network), Klassifizierungssubnetz und Regressionssubnetz [75]. Das Rückgrat bildet das Basisnetzwerk, welches die Merkmale des Eingabebildes, wie im vorherigen Abschnitt dargestellt, mittels *Convolution* und *Pooling Operation* extrahiert. Hierfür können verschiedenste Netzwerke zum Einsatz kommen. Im Falle dieser Arbeit bildete ein Netzwerk, basierend auf der ResNet Architektur, das Rückgrat Element [77].

Das *Feature Pyramid Network* (FPN) dient der Verbindung von Schichten unterschiedlicher Tiefe innerhalb des Netzwerks. Diese weisen je nach Position innerhalb des Netzwerks unterschiedliche Vor- und Nachteile auf. So können in tieferen Netzwerkschichten durch ein hohes Maß an Abstrahierung komplexere Bildmerkmale detektiert werden. Jedoch gehen somit Daten über die genaue Position der entsprechenden Merkmale verloren. Darüber liegende Schichten sind für die Objekterkennung von geringerer Bedeutung, weisen jedoch eine deutlich höhere Auflösung auf. Durch die Verbindung dieser Schichten können die jeweils

fehlenden Informationen ergänzt und somit bessere Ergebnisse bei der Objekterkennung erzielt werden.

Angeschlossen an das FPN sind die Subnetze. Hier findet die Zuordnung zu den vordefinierten Objektklassen (z.B. Gelenke oder Erosionsgrade) statt.

1.5 Ziele der Arbeit

Ein wesentlicher Bestandteil der diagnostischen Differenzierung der unterschiedlichen Arthritiden des rheumatischen Formenkreises sowie deren Früherkennung und Therapieplanung ist die bildgebende Beurteilung erosiver Gelenkveränderungen.

Expertise und Erfahrung im Umgang mit diesen komplexen Pathologien sind dabei von enormer Bedeutung, unterscheiden sich jedoch je nach Zentrum deutlich, woraus häufig eine verzögerte Diagnosestellung und ein konsekutiv verspäteter Behandlungsbeginn resultieren [78]. Zudem müssen alle neuen Wirkstoffe und Medikamente in der Rheumatologie neben der Symptomkontrolle auch beweisen, dass sie in der Lage sind, die radiologische Progression zu stoppen und damit den unwiderruflichen destruktiven Prozess und die drohende Funktionseinschränkung aufhalten zu können. Die Anfertigung, Kontrolle und Auswertung konventioneller Röntgenbilder sind daher weiterhin fester Bestandteil aller zulassungsrelevanten klinischen Studien, um dies auch den zulassenden Behörden gegenüber beweisen zu können und stellen in allen klinischen Studienprogrammen einen erheblichen Zeit- und Kostenfaktor dar.

Die Auswertung der bildgebenden Methoden stellt auch heute noch selbst erfahrene Radiologen vor Schwierigkeiten. Dies liegt zum einem an der Vielzahl der potenziell betroffenen Gelenkregionen, zum anderen an dem hohen zeitlichen Aufwand, welcher mit einer gründlichen Begutachtung der zahlreichen Regionen einhergeht. Eine Optimierung der Auswertealgorithmen bei konstanter Qualität ist daher ein bedeutendes Ziel der Forschung.

Automatisierte Methoden der Bilddatenanalyse können in diesem Zusammenhang einen großen Beitrag zur Diagnostik der Erkrankungen des rheumatischen Formenkreises und deren differentialdiagnostischer Differenzierung leisten und diagnostische Blindbereiche schließen.

Die Fortschritte im Bereich der Verfahren der Künstlichen Intelligenz im Allgemeinen und der Anwendung von *Deep-Learning* Technologien im Speziellen haben neue Wege in der automatisierten Verarbeitung medizinischer Bilddatensätze erschlossen [44, 45, 48, 50].

In diesem Zusammenhang konnte bereits gezeigt werden, dass die zum Einsatz kommenden künstlichen neuronalen Netzwerke in der Diagnosestellung

gleichwertig, zum Teil sogar überlegen im Vergleich zu geschulten Experten sind [46].

Die Diagnosen weisen darüber hinaus einen hohen Grad an Standardisierung und Reliabilität auf [46]. Erste Studien legen nahe, dass CNNs zur Evaluierung erosiver Veränderungen im Rahmen der rheumatoiden Arthritis auf Grundlage von Röntgenbildern herangezogen werden können [47, 51]. Diese Studien sind allerdings durch unzureichende Zahlen von Trainingsdatensätzen, mangelnde differentialdiagnostische Aufarbeitung und fehlende Objektivierung der Krankheitsschwere bzw. -aktivität gekennzeichnet.

Das Ziel der Arbeit war folglich, ein CNN zu entwickeln, welches zur reliablen Detektion von erosiven Veränderungen bei Patient:innen mit RA-, PsA- sowie sonstigen entzündlichen Arthritiden in der Lage ist.

2 Material und Methoden

2.1 Datenakquise

Aus den Datenbanken des Instituts für Diagnostische und Interventionelle Radiologie sowie der Klinik für Rheumatologie und des Hiller Forschungszentrums für Rheumatologie des Universitätsklinikums Düsseldorf wurden aus einer großen Anzahl hoch qualitativer und standardisierter Röntgenaufnahmen der Hände, die Aufnahmen von 208 gesichert an RA bzw. PsA erkrankten Patient:innen durch eine Abfrage über das docuMed® (klinikinternes KIS) System gesammelt. Davon konnten 89 Patient:innen aufgrund qualitativer Einschränkung oder fehlender Bilddatensätze nicht einbezogen werden.

Auswahlkriterien für die Patient:innen waren eine gesicherte, seropositive (Nachweis von CCP-AK und/oder Rheumafaktoren) RA bzw. PsA und mindestens eine Erosion. Eine Erosion wurde hierbei nach Larsen als eine Unterbrechung der Kortikalis um mindestens 1 mm definiert [79].

Zur Sicherung der Diagnose dienten sowohl die klinische Untersuchung durch einen Rheumatologen als auch laborchemische und fachärztliche radiologische Befunde. Es wurden sowohl Patient:innen im frühen Krankheitsstadium mit einer Erkrankungsdauer von unter 2 Jahren als auch etablierte Erkrankungen mit einer Erkrankungsdauer von über 2 Jahren und teils ausgeprägter erosiver Erkrankung einbezogen.

Der Datensatz enthielt schlussendlich 119 Patient:innen mit insgesamt 300 Röntgenstandardaufnahmen beider Hände in a.p. Projektion und insgesamt 9600 Gelenkregionen.

Von den ausgewählten Patient:innen waren 41 männlich und 78 weiblich mit einem Altersdurchschnitt von $55,6 \pm 12,2$ Jahre. Die Altersspanne lag bei 23 - 87 Jahre.

Die Auswahl der an RD's erkrankten Patient:innen wurde zur weiteren Verarbeitung pseudonymisiert und tabellarisch in Microsoft® Excel®, nach ihrer UKD spezifischen Patienten-ID sortiert, aufgelistet.

Die Bilder wurden aus dem UKD internen *Picture Archiving and Communication System* (PACS) heruntergeladen und lagen als DICOM (Digital Imaging and Communications in Medicine) 12-Bit-Graustufenbilder vor. Die Bildauflösung reichte von 1572 x 1572 bis 2738 x 2738 Pixel.

2.2 Scoring nach Sharp van der Heijde

Im nächsten Schritt wurden alle Röntgenaufnahmen von mir mithilfe des Sharp/van der Heijde Scores (SvdH) im Hinblick auf die Gelenkspaltverschmälerung und Erosionsausprägung evaluiert.

Die Auswertung fand auf einem kalibrierten Befundungsmonitor des Instituts für Diagnostische und Interventionelle Radiologie des Universitätsklinikums Düsseldorf statt.

Der SvdH Score ist ein in einer Vielzahl von Studien angewendeter und standardisierter Auswertungsscore für konventionelle Röntgenbilder, welcher eine hohe Reliabilität bei der Bewertung erosiver Veränderungen aufweist [80].

Der SvdH Score stellt eine Modifikation der 1971 vorgestellten und 1985 modifizierten Methode von Sharp et al. dar [81, 82].

Die Gelenkspaltverschmälerung und Erosionsausprägung werden dabei unabhängig voneinander für jedes Gelenk bzw. jede Gelenkfläche bestimmt und anschließend in einem Gesamtwert zusammengefasst.

Im Regelfall inkludiert der Score eine Mitbewertung der Füße, jedoch wurde im Rahmen dieser Arbeit der Fokus auf die alleinige Analyse der Hände und deren Gelenkregionen gelegt.

2.2.1 Erosionsscore

Der Erosions-Score umfasst die Bewertung von jeweils 16 Gelenken der Hand und Handwurzel beidseits. Darunter die 5 MCP und 4 PIP-Gelenke, das Interphalangealgelenk des Daumens (IP), das erste Carpometacarpal-Gelenk (CMC), den distalen Radius, die distale Ulna sowie die Gelenkflächen des Os trapezium-trapezoideum, Os scaphoideum und Os lunatum. Die Erosionsausprägung wird mit minimal 0 bis maximal 5 Punkten bewertet. Die Auswahl des entsprechenden Wertes richtet sich nach der Größe und Anzahl der Erosionen.

Insgesamt ist ein Erosionsscore von 0-160 für beide Hände möglich. Durch diese differenzierte Betrachtung lassen sich sowohl die gelenkspezifische als auch die gesamte Erosionsausprägung der Patient:innen mit dem Score abbilden. So können, insbesondere im Zeitalter moderner Therapieoptionen und Interventionen, minimale Veränderungen von erosiven Prozessen im Verlauf detektiert werden.

Tabelle 3: Erosions-Score nach Sharp van der Heijde

Erosions-Score	Erosive Veränderungen
0	normal
1	1 kleine Erosion
2	1 längere oder 2 kleine Erosionen
3	die Mittellinie überschreitende oder 3 kleine Erosionen oder 2 kleine + 1 längere Erosion
4	4 kleine oder 2 längere oder 1 die Mittellinie überschreitende + 1 kleine Erosion
5	Summe der Erosionen 5 oder mehr



Abbildung 3: Beispielgelenke mit korrespondierendem Erosionsscore nach Sharp van der Heijde aus dem genutzten Datensatz

2.2.2 Gelenkspaltverschmälerungs-Score

Die Gelenkspaltverschmälerung (engl.: Joint Space Narrowing, kurz: JSN) wird nach SvdH an jeweils 15 Regionen beider Hände bestimmt.

Darunter die Gelenkspalten der 5 MCP und 4 PIP-Gelenke, der CMC-Gelenke des dritten bis fünften Phalangen sowie jeweils zwischen Os trapezium-trapezoideum, scaphoideum-capitatum und radiocarpal [83].

Die Bewertung richtet sich nach dem Ausmaß der Gelenkspaltverschmälerung und kann mit minimal 0 bis maximal 4 Punkten bewertet werden.

Erweiternd zu der ursprünglichen Methode nach Sharp, werden Subluxation und Luxation in die Gelenkspaltbewertung miteinbezogen.

Folglich ist ein JSN-Score von 0-120 und in Zusammenschau mit dem Erosionsscore ein Gesamtscore für beide Hände von 280 möglich.

Tabelle 4: JSN-Score nach Sharp van der Heijde

JSN-Score	Ausprägung der Gelenkspaltverschmälerung
0	normal
1	fokal oder nicht sicher beurteilbar/mögliche Verschmälerung
2	diffuse Gelenkspaltverschmälerung um weniger als 50%
3	diffuse Gelenkspaltverschmälerung um mehr als 50% oder Subluxation
4	knöchernen Ankylose oder komplette Luxation

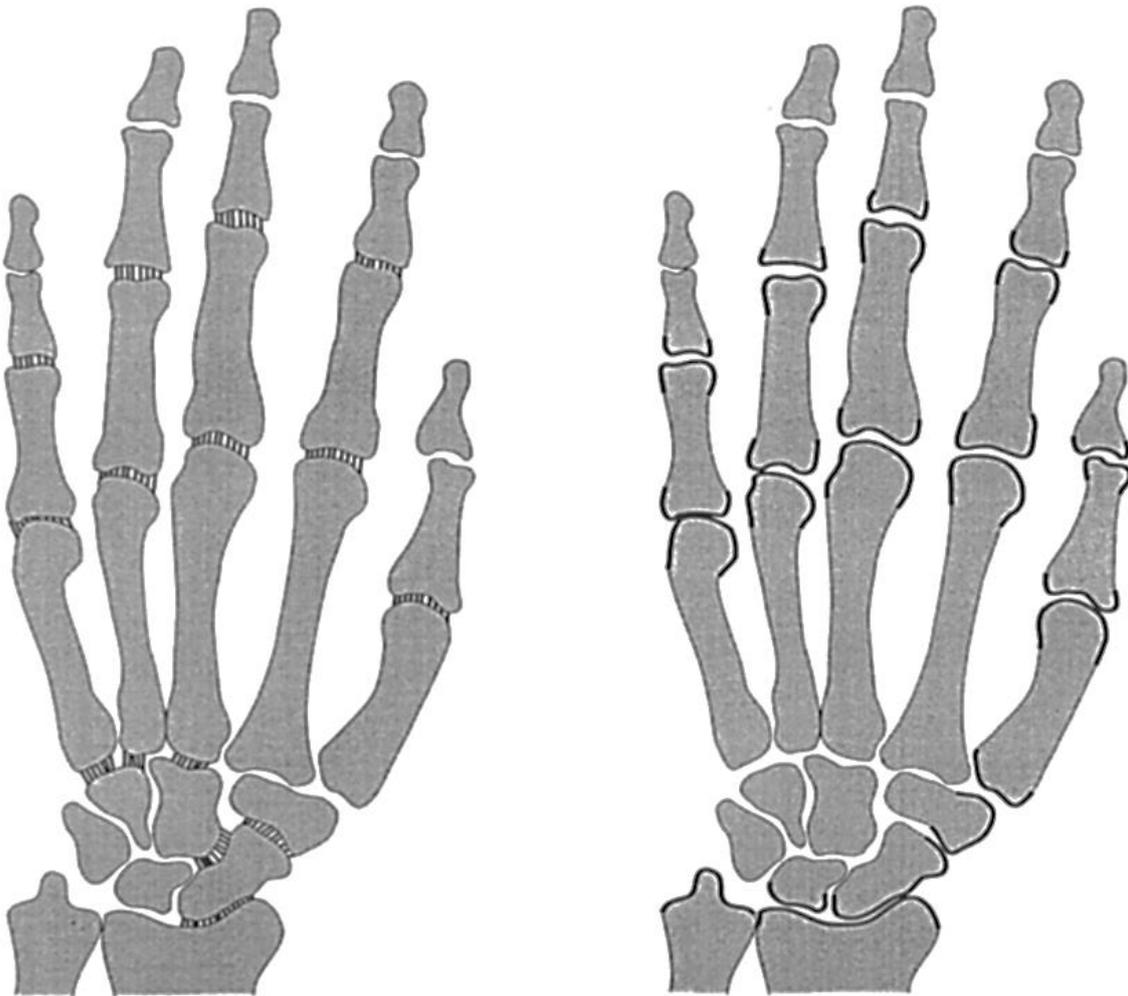


Abbildung 4: Bewertete Gelenke und Gelenkflächen im Rahmen des Sharp/van der Heijde Scores (modifiziert nach Rau et al, 2007):

Gelenkspaltverschmälerung (links), Erosionen (rechts)

2.3 Software und Hardware

Um eine simultane Lokalisation und Erosionsklassifizierung zu ermöglichen, wurde ein RetinaNet als Grundlage für die Netzwerkerstellung verwendet.

Abbildung 7 zeigt schematisch den Aufbau des verwendeten Netzwerkes.

Es setzt sich zusammen aus einem *Residual Neural Network* (ResNet) (a) und drei Subnetzen (c-e). Das ResNet bildet das Rückgrat des Systems. Hier findet die klassische Faltungsoperation statt. Die Bildinformationen werden extrahiert und komprimiert. Im Anschluss werden die auf diese Weise gewonnenen Bildinformationen im *Feature Pyramid Network* verarbeitet.

Das Gelenk Klassifizierungs-Subnetz ordnet die erkannten Interessensregionen einem Gelenk (z.B. MCP II rechts) zu. Das SvdH-Score Klassifizierungs-Subnetz bestimmt den Erosions-Score der erkannten Gelenkregion.

Alle Trainingsläufe wurden auf den Servern des Universitätsklinikum Düsseldorf ausgeführt. Der genutzte Server war ausgestattet mit einem Intel® Xeon Gold 6242R Prozessor (Intel Corporation, Santa Clara, CA, USA), 376GB

Hauptspeicher sowie vier NVIDIA® RTX 3090 Grafikkarten (NVIDIA, Santa Clara, CA, USA).

Die Programmierung und das Training des Netzwerkes wurden durch Mitarbeiter des Instituts für Diagnostische und Interventionelle Radiologie des UKD vorgenommen.

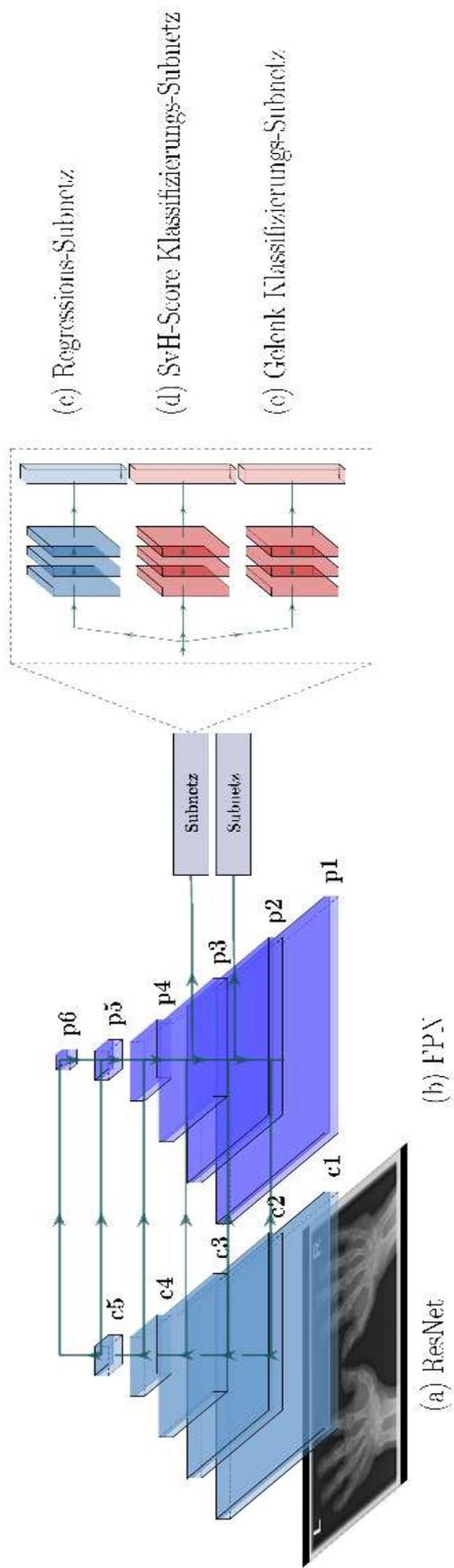


Abbildung 5: Schematischer Aufbau des RetinaNet zur Lokalisation und Klassifizierung der Röntgenbilder der Hände [1].

- (a) Residual Neural Network (ResNet)
- (b) Feature Pyramid Network (FPN)
- (c) Regressions Subnetz
- (d) SvH-Score Klassifizierungs-Subnetz
- (e) Gelenk Klassifikations-Subnetz

2.4 Präprozessierung und Augmentation

Die Qualität der eingegebenen Daten ist von großer Bedeutung bei der Erstellung eines effizienten neuronalen Netzwerks. Aus diesem Grund wurde eine Vorverarbeitung der Bilddatensätze durchgeführt.

Je nach zugrundeliegenden DICOM Bildinformationsdaten wurden die Röntgenbilder gefenstert und zwischen 0 und 4096 skaliert.

Außerdem erfolgte eine Z-Score-Standardisierung der Pixelwerte. Die Angleichung all dieser Bildparameter ist wichtig, um eine effiziente Verarbeitung der Bilder durch das Netzwerk zu ermöglichen.

Die Leistung eines neuronalen Netzwerks ist außerdem stark abhängig von der Menge an zur Verfügung stehenden Daten. Insbesondere im Falle von medizinischen Fragestellungen sind die Datensätze durch nur geringe Patientenzahlen in ihrer Größe beschränkt. Eine Vergrößerung des Datensatzes (engl. Augmentation) ist daher ein wichtiges Instrument im Bereich des maschinellen Lernens.

Zur Augmentation des Datensatzes wurde die Python Bibliothek „imgaug“ verwendet [84]

Die Software ermöglicht eine Anpassung festgelegter Bildparameter in einem vorgegebenen Rahmen. So wurden die Bilder in der x- und y-Achse um bis zu 5% der Bildgröße gestaucht, zwischen +/- 15 Grad rotiert und mit dem Faktor 0,5-2 skaliert. Der Bildkontrast wurde zudem zwischen 0,5 -1,5 variiert. Auch Schärfe und Belichtung der Bilder wurden verändert und folglich der Datensatz vergrößert.

Beispielbild und erzeugte Datenerweiterungen

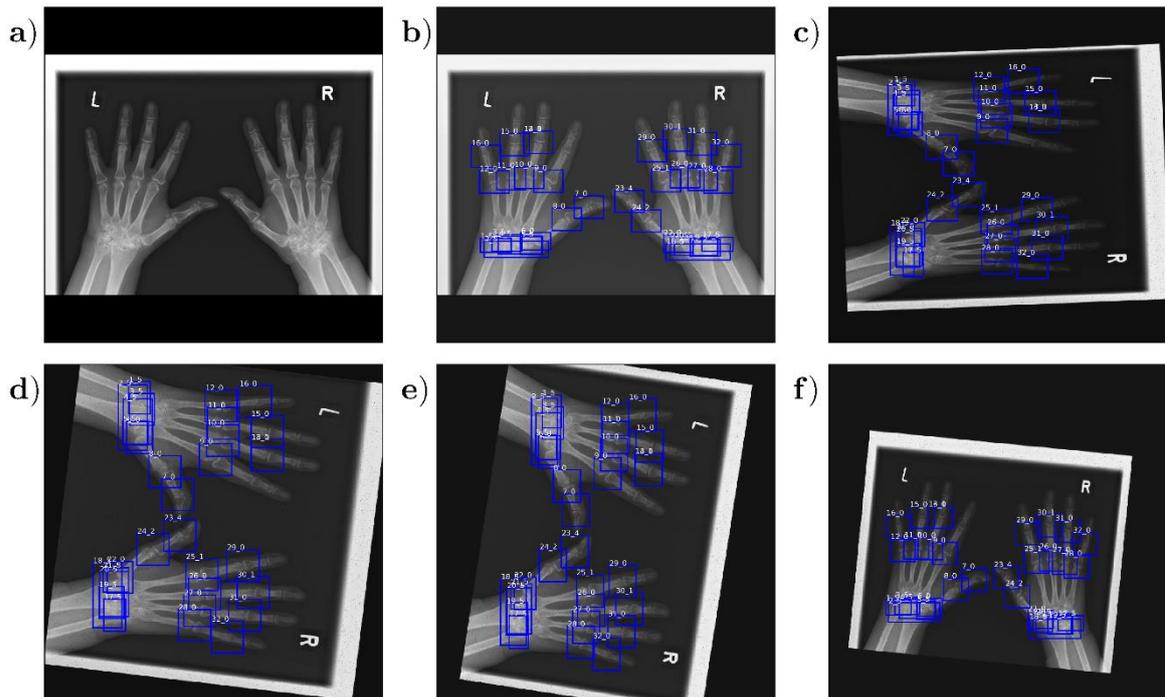


Abbildung 6: Repräsentative Darstellung der Augmentation [1]

Die Abbildung zeigt sowohl das ursprüngliche Röntgenbild in unveränderter Version (a) als auch mit markierten Interessensregionen in blau (b) sowie vier Beispiele für eine Augmentation (c-f). Zudem sind durch einen Unterstrich getrennte Zahlen über den Interessensregionen zu sehen. Die erste Zahl bezeichnet die spezifische Nummer des Gelenks. Diese wurden nach einer festgelegten Reihenfolge von 1-32 durchnummeriert. Die zweite Zahl entspricht dem durch das Netzwerk zugeordneten Erosions-Score des Gelenks.

2.5 Gelenklokalisierung und Erosionsbewertung durch das CNN

Die Analyse der Handröntgenbilder bestand aus der automatisierten Detektion und Erosionsbewertung der identifizierten Gelenkregionen.

Zu diesem Zweck wurde die Größe der Gelenkregionen bei 20 Patienten mittels ITK-SNAP®, einer Softwareanwendung zur Darstellung und Segmentation medizinischer Bilddatensätze [85], in Form von x- und y-Koordinaten markiert.

Die resultierenden Interessensregionen sowie die manuell bestimmten SvdH Scores dienten im weiteren Verlauf als Grundwahrheit.

Es folgte eine Aufteilung des Datensatzes in 3 Teile: Training (70%, Anzahl der Patienten = 83, Anzahl der Bilder = 231), Validierung (10%; n = 12, Anzahl der Bilder = 45) und Test (20%, n = 24, Anzahl der Bilder = 24)

Die Modelle wurden basierend auf den Trainingsdaten über 100 Epochen trainiert. Grundlage für die Auswahl des Netzwerks mit der höchsten Genauigkeit war die

Leistung im Validierungsdatensatz. Das optimale Netzwerk wurde ermittelt und die Leistung auf den noch ungesehenen Bildern des Testdatensatzes überprüft.

2.6 Bewertungskriterien

Um die Leistung der unterschiedlichen Netzwerkmodelle zu vergleichen, wurden verschiedene Parameter herangezogen.

Erstens wurde die Gesamtgenauigkeit bestimmt. Diese bildet sich aus dem Anteil korrekt identifizierter Gelenke an der Gesamtheit aller vorhandenen Gelenke.

Um als korrekt identifiziertes Gelenk gewertet zu werden, muss ein Gelenk sowohl an der richtigen Stelle im Röntgenbild lokalisiert als auch dem korrekten Erosionsscore zugeordnet worden sein.

Zweitens wurde die mittlere Durchschnittsgenauigkeit (engl. mean average precision, kurz: mAP) ermittelt. Diese dient als wesentlicher Indikator für die Netzwerkleistung. Hierfür wird die Genauigkeit des Netzwerks für jeden Erosionsgrad einzeln bestimmt und anschließend der Durchschnitt gebildet. Im Gegensatz zur Gesamtgenauigkeit wird die Netzwerkleistung somit akkurater wiedergegeben, da ein Klassenungleichgewicht geringere Auswirkungen auf die Ergebnisse zur Folge hat.

Drittens wurde die Reproduzierbarkeit der Erosionsbewertung durch den menschlichen Bewerter überprüft. Zu diesem Zweck wurden zwei Bewertungszeitpunkte definiert: Zum Zeitpunkt 1 wurde die erstmalige Analyse der Röntgenbilder vorgenommen. Zum Zeitpunkt 2 (nach 6 Monaten) wurde erneut eine repräsentative Schnittmenge des Datensatzes auf Erosionsausprägung untersucht. Auf diese Weise konnte die Abweichung des Bewerter ermittelt werden.

2.7 Statistische Analyse

Die statistischen Analysen wurden mit Hilfe der Open-Source-Software "SciPy" durchgeführt.

Zur Berechnung der statistischen Beziehungen zwischen den Erosions-Scores des menschlichen Befunders und des Netzwerks wurde der Kendall-Tau-Korrelationskoeffizient (τ) bestimmt. Die Tau-Effektgröße wurde nach Cohen al. als gering (0,1-0,3), mittel (0,3-0,5) und stark ($>0,5$) eingestuft [86].

Um eine Kumulierung des Alpha-Fehlers zu vermeiden, wurde das Signifikanzniveau von $p \leq 0,05$ nach der Bonferroni-Methode auf $p \leq 0,0083$ angepasst.

2.8 Ethikvotum

Für die Studie „Bildgebende Differenzierung von rheumatischen Erkrankungen mittels Deep Learning Algorithmen - Vom Röntgenbild zur MRT Studie“ (Studien-Nr.: 2019-648-andere Forschung erstvotierend) liegt ein positives Ethikvotum vom 10.12.2019 vor.

3 Ergebnisse

3.1 Patientenmerkmale

Nach zufälliger Zuordnung der Patient:innen auf die drei Datensätze (Training, Validierung und Test) ergaben sich 3 in ihren medizinischen und demografischen Merkmalen annähernd gleiche Gruppen.

Die Patient:innen im Testdatensatz hatten ein Durchschnittsalter von 57 ± 12 Jahren (27 - 87) und einen mittleren Gesamterosionsscore von 55 ± 30 (32 - 179).

Tabelle 5: Überblick der Patient:innen Charakteristika.

Bei Vorhandensein von mehr als einem Bild pro Patient wurden das Alter und der Score des ersten Röntgenbildes verwendet.

Parameter	Training	Validation	Test
Alter	55 ± 13 (23 – 87)	56 ± 7 (29 – 82)	57 ± 12 (27 – 87)
männlich / weiblich	29 \ 54	4 \ 8	6 \ 18
Summe der Erosionen	53 ± 25 (32 – 179)	46 ± 8 (32 – 101)	55 ± 30 (32 – 179)
Durchschnittlicher Erosionsscore	1.65 ± 0.77 (1.0 – 5.59)	1.43 ± 0.25 (1.0 – 4.3)	1.72 ± 0.95 (1 – 5.59)
Minimaler Erosionsscore	1.0 ± 0 (1 - 1)	1.0 ± 0.0 (1 – 1)	1.0 ± 0.0 (1-1)
Maximaler Erosionsscore	4.2 ± 1.87 (1 – 6)	$3.2 \pm - 1.7$ (1 – 6)	4.6 ± 1.6 (1 – 6)

Die unterschiedlichen Erosionsgrade waren im Testdatensatz analog zum gesamten Patientenkollektiv in folgender Verteilung vertreten:

Bei einer Gesamtanzahl von 768 Gelenken im Testdatensatz machten gesunde Gelenke ohne Erosion 75,26% der Gelenke aus. 6.9% der Gelenke wiesen einen Erosionsscore von 1, 4.56% einen Erosionsscore von 2, 3.52% einen Erosionsscore von 3, 3.13% einen Erosionsscore von 4 und 6.64% einen Erosionsscore von 5 auf.

Tabelle 6: Übersicht über die Verteilung der Erosionsgrade nach dem SvH-Erosionsscore auf der Grundlage der 768 Gelenke im Testdatensatz

Erosionscore nach Sharp v.d. Hejde	Anzahl Gelenke	Prozentualer Anteil
0	578	75.26%
1	53	6.90%
2	35	4.56%
3	27	3.52%
4	24	3.13%
5	51	6.64%

3.2 Übereinstimmung von automatischer und manueller Gelenkbewertung

Das Netzwerkmodell mit der höchsten Leistung erreichte eine hohe Übereinstimmung von automatisierter und manueller Gelenkbewertung mit einer Genauigkeit von 94% und einer mAP von 0.81. Die höchste Klassengenauigkeit lag bei circa 99% und die niedrigste bei etwa 48%.

Die Intra-Rater-Reliabilität lag nach 6 Monaten über alle Klassen bei 88,5% mit einer mAP von 0.79.

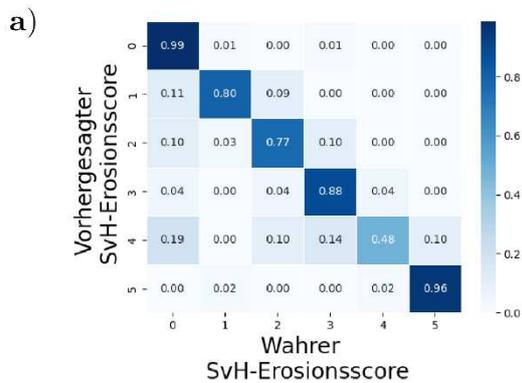
Darüber hinaus konnte eine signifikante Korrelation der automatisierten und manuellen Erosionsbewertung sowohl für die einzelnen Gelenke ($\tau = 0,92$, $p < 0,001$) als auch für den Summenscore ($\tau = 0,88$, $p < 0,001$) nachgewiesen werden. Auch die beiden Bewertungszeitpunkte des menschlichen Befunders zeigten eine signifikante Korrelation bezüglich der einzelnen Gelenkscores ($\tau = 0.90$, $p < 0.001$) und des Summenscores ($\tau = 0.92$, $p < 0.001$).

Die Ergebnisse sind in Abbildung 8 in Form einer Konfusionsmatrix dargestellt.

Sowohl der menschliche Befunder als auch das Netzwerk klassifizierten die Gelenke mit einem Score von 0 bzw. 5 mit hoher Konstanz.

Das RetinaNet tendiert jedoch stärker dazu, Gelenke mit einem Score von 0 zu bewerten, wohingegen der menschliche Befunder zum Zeitpunkt 2 ausschließlich um ± 1 Score von der vorherigen Bewertung abwich.

Vorhersage RetinaNet
vs. Grundwahrheit



Zweite Vorhersage
vs. Grundwahrheit

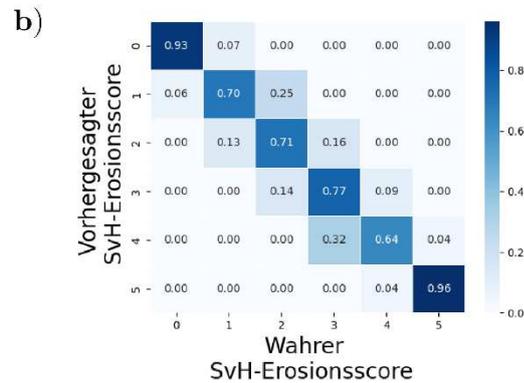


Abbildung 7: Ergebnisse der Erosionsbewertung in Form einer Konfusionsmatrix [1]:

Die X-Achse zeigt die Grundwahrheit, also die Erosionsbewertung durch den menschlichen Befunder zum Zeitpunkt 1. Die Y-Achse repräsentiert die Erosionsbewertung durch das Netzwerk (a) beziehungsweise die Erosionsbewertung durch den menschlichen Befunder zum Zeitpunkt 2 (b).

3.3 Zeitaufwand

Eine manuelle Bewertung aller Gelenkregionen durch einen erfahrenen Radiologen benötigt inklusive Dokumentation durchschnittlich etwa 9 Minuten ($\pm 13,23$ min). Der durchschnittliche Zeitbedarf der Algorithmen basierten Analyse eines einzelnen Röntgenbildes lag hingegen bei unter 1 Sekunde auf einem spezialisierten Rechner. Auf einem innerklinischem Standardrechner eines Radiologen lag die Analysedauer bei etwa 5 Sekunden, was einer Zeitersparnis von über 99% entspricht.

4 Diskussion

Das Ziel der vorliegenden Arbeit war die Entwicklung einer automatisierten Methode zur Analyse erosionsspezifischer Veränderung bei RD's. Hierzu wurde auf Basis aktueller Methoden des maschinellen Lernens erfolgreich ein Netzwerk entwickelt, welches hohe Genauigkeiten bei der Gelenklokalisierung und Erosionsbewertung aufweist. Auch anatomisch komplexe Regionen wie die Handwurzel wurden zuverlässig und mit hoher Präzision detektiert und analysiert. Darüber hinaus erreichte das Netzwerk eine hohe Verarbeitungsgeschwindigkeit bei geringem Hardwarebedarf und stellt damit ein gut implementierbares Werkzeug für die klinische Praxis dar.

4.1 Maschinelles Lernen in der Rheumatologie

Die Techniken des maschinellen Lernens sind in den vergangenen Jahren immer weiter in den Fokus der Forschung getreten. Über alle Fachrichtungen in der Medizin hinweg zeigt sich enormes Interesse an *Deep Learning* Ansätzen zur Automatisierung und Prozessoptimierung der klinischen Diagnostik und Therapie. Dies spiegelt sich unter anderem im rapiden Anstieg der Anzahl an Publikationen mit Fokus auf *Deep Learning* und *Convolutional Neural Networks* im medizinischen Kontext wieder [87].

Innovative Ansätze zur Algorithmen basierten Analyse medizinischer Bilddatensätze besitzen in diesem Zusammenhang besonders großes Potential, klinische Abläufe zu beschleunigen und zugleich personelle Ressourcen zu schonen.

In der Rheumatologie ist die Bildgebung mittels CT, MRT und konventionellen Röntgenuntersuchungen ein entscheidendes Element der diagnostischen Entscheidungsfindung und Therapiekontrolle. Eine Implementierung dieser Technologien ist daher besonders relevant.

Die hohe Komplexität der mit RD's assoziierten, bildmorphologischen Krankheitszeichen macht jedoch spezifisch zugeschnittene *Deep Learning* Algorithmen erforderlich, um eine verlässliche klinische Anwendung zu ermöglichen.

Bereits 2002 wurde in einer Studie von Duryea et al. ein frühe Version eines neuronalen Netzwerk erfolgreich zur Detektion häufig involvierter Gelenkregionen bei Osteoarthritis eingesetzt [88]. Sie erreichten eine Erfolgsrate von 81-99% bei

der automatisierten Identifikation der carpometacarpal (CMC), radiocarpal (RC) und scaphocapital (SC) Gelenke. Die Anzahl der detektierten Gelenkregionen war jedoch nicht ausreichend, um einen sinnvollen klinischen Nutzen zu erbringen. Es erfolgte zudem auch keine Beurteilung der erosiven Veränderungen, welches für eine klinische Anwendung von hoher Bedeutung ist.

Mit der Vorstellung des ResNet durch He et al. im Jahr 2016 [77], wurde die Grundlage für die Erstellung von Netzwerkmodellen geschaffen, welche die Analyse von komplexeren Strukturen, wie z.B. erosiven Veränderungen der Gelenkregionen, deutlich verbesserten. Aufbauend auf diesem Konzept konnten unterschiedliche Ansätze zur Analyse erosionsspezifischer Veränderungen entwickelt werden.

Erste Arbeiten wiesen jedoch vor allem Einschränkungen im Hinblick auf die klinische Anwendbarkeit auf. So beschränkten sich sowohl die Arbeiten von Murakami et al. als auch Üreten et al. auf die bloße Detektion von Erosionen, ohne deren Ausprägung näher zu quantifizieren [51, 89]. Dies ist jedoch von hoher klinischer Relevanz, da die Erosionsausprägung und Entwicklung im Krankheitsverlauf eng mit der Funktionskapazität von Patient:innen mit RA zusammenhängt [90] und zudem für die Auswertung im Rahmen von klinischen Studien entscheidend ist.

Zudem bestand bei einer Sensitivität von 78,9% (Murakami et al.) bzw. 73,3% (Üreten et al.) noch Potential für weitere Optimierung.

Die Arbeiten von Hirano et al. und Wang et al. präsentierten schließlich die erfolgreiche Implementation der SvdH-Methode in Kombination mit Deep Learning Algorithmen [91, 92]. Auf diese Weise wurde die Abgrenzung verschiedener Schweregrade von Erosionsausprägung verbessert.

In den Arbeiten wurden darüber hinaus hohe Genauigkeiten in der Gelenklokalisierung und Erosionsbewertung der PIP und MCP-Gelenke erreicht. Eine Abgrenzung der Intercarpalgelenke stellte jedoch ein großes Hindernis dar, da diese aufgrund ihrer anatomischen Konfiguration von den entsprechenden Netzwerken oft gar nicht oder nur fehlerhaft erkannt wurden.

4.2 Vorteile des RetinaNet

Mit dem auf dem RetinaNet basierenden Ansatz wählten wir für diese Arbeit eine Netzwerkstruktur, welche hohe Genauigkeit mit verbesserter Analyse komplexer Bildbereiche kombiniert. So konnten wir nicht nur eine hohe Genauigkeit in der Detektion von PIP/IP und MCP-Gelenken, sondern auch der Intercapalgelenke erreichen. Dies gewährleistet eine größere Übereinstimmung mit dem im klinischen Setting angewandtem SvdH Score.

Ein weiterer Vorteil des RetinaNet gegenüber anderen Netzwerktypen ist die Möglichkeit der visuellen Repräsentation der für die Entscheidung relevanten Bildbereiche. Diese können als Interessensbereiche auf den Röntgenbildern dargestellt werden und so im weiteren Verlauf durch einen Experten überprüft und bestätigt werden.

4.3 Besonderheiten und Limitationen

Trotz einer hohen Übereinstimmung der maschinellen und menschlichen Erosionsbewertung bei hoher Reliabilität, sind Limitationen der Arbeit zu nennen. Zum einen stehen aktuell noch keine standardisierten und institutionsübergreifenden Datensätze zur Verfügung. Folglich war die Anzahl der verfügbaren Patient:innen limitiert.

Des Weiteren wurden in dieser Arbeit die Effekte verschiedener klinikinterner Plattformen und Hersteller von Hardware und Software auf die Ergebnisse nicht berücksichtigt.

Obwohl Röntgenbilder anders als MRT-Aufnahmen ein hohes Maß an Standardisierung über die Institutionengrenzen hinweg aufweisen, sind weitere Untersuchungen notwendig, um eine Anwendbarkeit des entwickelten Netzwerkmodells auf Datensätze anderer Institute zu bestätigen.

4.4 Ausblick

Das in dieser Arbeit vorgestellte Netzwerk für die automatisierte Analyse erosionsspezifischer, röntgenologischer Krankheitszeichen bei RD's bietet aufgrund seiner hohen Genauigkeit eine solide Grundlage für die Entwicklung einer Software zur Implementation in den klinischen Alltag. Bereits zu diesem frühen Zeitpunkt kann das Netzwerk unerfahrene Radiolog:innen bei der Diagnosestellung unterstützen. Es sind jedoch noch einige Ergänzungen erforderlich, damit diese Technologie ihr volles diagnostisches Potential ausschöpfen kann.

Ein wesentlicher Aspekt, der zur Weiterentwicklung der Software beitragen kann, ist die Erweiterung um die automatisierte Analyse der Gelenkspaltverschmälerung. Das für diesen Aspekt notwendige Scoring nach SvdH wurde für den jetzigen Datensatz bereits durchgeführt und kann im Verlauf ebenfalls trainiert, getestet und zum Algorithmus hinzugefügt werden.

Die in dieser Arbeit exkludierte Bewertung der Fußgelenke im SvdH Score kann im Verlauf ebenso ergänzt werden. Nur so können die erosiven Veränderungen in Gänze erfasst und adäquat bewertet werden.

Aber auch nicht bildgebende Informationen (z.B. Labor, ACR-Kriterien) sind von hoher Relevanz bei der Erstdiagnose und Verlaufskontrolle von RD's. Das langfristige Ziel ist es, ein Netzwerk zu entwickeln, welches unter Berücksichtigung aller bildgebender, laborchemischer und klinischer Informationen eine differenzierte und sicherere Aussage über die Krankheitswahrscheinlichkeit und den Krankheitsverlauf treffen kann.

Das größte Hindernis bei der Entwicklung von Deep Learning Algorithmen ist jedoch die Größe des zur Verfügung stehenden Datensatzes. Eine nationale oder internationale Kooperation und Datenakquise unter standardisierten Gesichtspunkten könnten einen enormen Beitrag zur Weiterentwicklung der bereits existierenden Netzwerke beitragen.

Danksagung

Ich bedanke mich für die liebevolle Unterstützung meiner Familie, die mich in meinem privaten oder beruflichen Werdegang stets unterstützt hat.

Mein besonderer Dank geht an meinen Doktorvater Priv.-Doz. Dr. Philipp Sewerin, der mir bei all meinen Fragen und Problemen zur Seite stand und mich über die Dauer meiner Promotion äußerst professionell betreut und begleitet hat.

Ebenfalls möchte ich mich bei den Mitarbeitern des Instituts für Diagnostische und Interventionelle Radiologie des Universitätsklinikums Düsseldorf für die produktive Zusammenarbeit bedanken sowie bei allen Ko-Autoren unserer Publikation.

Literaturverzeichnis

1. Radke, K.L., et al., *Adaptive IoU Thresholding for Improving Small Object Detection: A Proof-of-Concept Study of Hand Erosions Classification of Patients with Rheumatic Arthritis on X-ray Images*. Diagnostics, 2023. **13**(1): p. 104.
2. Okada, Y., et al., *Genetics of rheumatoid arthritis: 2018 status*. Ann Rheum Dis, 2019. **78**(4): p. 446-453.
3. Stolt, P., et al., *Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases*. Ann Rheum Dis, 2003. **62**(9): p. 835-41.
4. Myasoedova, E., et al., *Epidemiology of rheumatoid arthritis: rheumatoid arthritis and mortality*. Curr Rheumatol Rep, 2010. **12**(5): p. 379-85.
5. Matcham, F., et al., *The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis*. Semin Arthritis Rheum, 2014. **44**(2): p. 123-30.
6. Scott, D.L., et al., *Long-term outcome of treating rheumatoid arthritis: results after 20 years*. Lancet, 1987. **1**(8542): p. 1108-11.
7. Doran, M.F., et al., *Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period*. Arthritis Rheum, 2002. **46**(3): p. 625-31.
8. Ruof, J., et al., *Costs of rheumatoid arthritis in Germany: a micro-costing approach based on healthcare payer's data sources*. Ann Rheum Dis, 2003. **62**(6): p. 544-9.
9. Arnett, F.C., et al., *The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis*. Arthritis Rheum, 1988. **31**(3): p. 315-24.
10. Bukhari, M., et al., *Erosions in inflammatory polyarthritis are symmetrical regardless of rheumatoid factor status: results from a primary care-based inception cohort of patients*. Rheumatology (Oxford), 2002. **41**(3): p. 246-52.
11. Schneider, M., et al., *[Management of early rheumatoid arthritis : Interdisciplinary guideline]*. Z Rheumatol, 2020. **79**(Suppl 1): p. 1-38.
12. Visser, H., et al., *How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis*. Arthritis Rheum, 2002. **46**(2): p. 357-65.
13. Aletaha, D., et al., *2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative*. Arthritis Rheum, 2010. **62**(9): p. 2569-81.
14. Ostergaard, M. and M. Boesen, *Imaging in rheumatoid arthritis: the role of magnetic resonance imaging and computed tomography*. Radiol Med, 2019. **124**(11): p. 1128-1141.
15. da Mota, L.M., et al., *Imaging diagnosis of early rheumatoid arthritis*. Rev Bras Reumatol, 2012. **52**(5): p. 757-66.
16. Colebatch, A.N., et al., *EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis*. Ann Rheum Dis, 2013. **72**(6): p. 804-14.
17. Narvaez Garcia, J.A., *[Evaluation through imaging of early rheumatoid arthritis]*. Reumatol Clin, 2010. **6**(2): p. 111-4.
18. Choy, E.H., et al., *Treating rheumatoid arthritis early with disease modifying drugs reduces joint damage: a randomised double blind trial of sulphasalazine vs diclofenac sodium*. Clin Exp Rheumatol, 2002. **20**(3): p. 351-8.
19. Mottonen, T., et al., *Delay to institution of therapy and induction of remission using single-drug or combination-disease-modifying antirheumatic drug therapy in early rheumatoid arthritis*. Arthritis Rheum, 2002. **46**(4): p. 894-8.
20. Finckh, A., et al., *Long-term impact of early treatment on radiographic progression in rheumatoid arthritis: A meta-analysis*. Arthritis Rheum, 2006. **55**(6): p. 864-72.
21. Astin, J.A., et al., *Psychological interventions for rheumatoid arthritis: a meta-analysis of randomized controlled trials*. Arthritis Rheum, 2002. **47**(3): p. 291-302.

22. Sewerin, P., et al., *Prevalence and incidence of psoriasis and psoriatic arthritis*. Ann Rheum Dis, 2019. **78**(2): p. 286-287.
23. Alinaghi, F., et al., *Prevalence of psoriatic arthritis in patients with psoriasis: A systematic review and meta-analysis of observational and clinical studies*. J Am Acad Dermatol, 2019. **80**(1): p. 251-265.e19.
24. Gladman, D.D., et al., *Psoriatic arthritis (PSA)--an analysis of 220 patients*. Q J Med, 1987. **62**(238): p. 127-41.
25. Dougados, M., et al., *The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy*. Arthritis Rheum, 1991. **34**(10): p. 1218-27.
26. Moll, J.M. and V. Wright, *Psoriatic arthritis*. Semin Arthritis Rheum, 1973. **3**(1): p. 55-78.
27. McGonagle, D., P.G. Conaghan, and P. Emery, *Psoriatic arthritis: a unified concept twenty years on*. Arthritis Rheum, 1999. **42**(6): p. 1080-6.
28. Fournie, B., et al., *Proposed classification criteria of psoriatic arthritis. A preliminary study in 260 patients*. Rev Rhum Engl Ed, 1999. **66**(10): p. 446-56.
29. Taylor, W., et al., *Classification criteria for psoriatic arthritis: development of new criteria from a large international study*. Arthritis Rheum, 2006. **54**(8): p. 2665-73.
30. Altmeyer, P. *Psoriasis arthropathica*. 2018 [cited 13.02.2021; Available from: <https://www.altmeyers.org/de/dermatologie/psoriasis-arthropathica-3341#authors>.
31. Tang, M.M., et al., *Quality of life and cost of illness in patients with psoriasis in Malaysia: a multicenter study*. Int J Dermatol, 2013. **52**(3): p. 314-22.
32. Nast, A., et al., *S3-Leitlinie zur Therapie der Psoriasis vulgaris*. 2017: AWMF online.
33. Harle, P., et al., *[Detection of psoriasis arthritis with the GEPARD patient questionnaire in a dermatologic outpatient setting]*. Z Rheumatol, 2010. **69**(2): p. 157-60, 162-3.
34. Harle, P., et al., *Sensitivity of the GEPARD Patient Questionnaire to Identify Psoriatic Arthritis in Patients with Psoriasis in Daily Practice: The GEPARD-Life Study*. Dermatology, 2016. **232**(5): p. 597-605.
35. Smolen, J.S., et al., *EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update*. Annals of the Rheumatic Diseases, 2022: p. ard-2022-223356.
36. Coates, L.C., et al., *Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA): updated treatment recommendations for psoriatic arthritis 2021*. Nature Reviews Rheumatology, 2022. **18**(8): p. 465-479.
37. Gossec, L., et al., *Management of psoriatic arthritis in 2016: a comparison of EULAR and GRAPPA recommendations*. Nat Rev Rheumatol, 2016. **12**(12): p. 743-750.
38. Gossec, L., et al., *European League Against Rheumatism (EULAR) recommendations for the management of psoriatic arthritis with pharmacological therapies: 2015 update*. Ann Rheum Dis, 2016. **75**(3): p. 499-510.
39. Coates, L.C., et al., *Group for Research and Assessment of Psoriasis and Psoriatic Arthritis 2015 Treatment Recommendations for Psoriatic Arthritis*. Arthritis Rheumatol, 2016. **68**(5): p. 1060-71.
40. Burkov, A., *The Hundred-Page Machine Learning Book*. 2019: Andriy Burkov.
41. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. Nature, 2015. **521**(7553): p. 436-44.
42. Ma, J., et al., *Deep neural nets as a method for quantitative structure-activity relationships*. J Chem Inf Model, 2015. **55**(2): p. 263-74.
43. Helmstaedter, M., et al., *Connectomic reconstruction of the inner plexiform layer in the mouse retina*. Nature, 2013. **500**(7461): p. 168-74.
44. Andersen, J.K.H., et al., *Neural networks for automatic scoring of arthritis disease activity on ultrasound images*. RMD Open, 2019. **5**(1): p. e000891.
45. Christensen, A.B.H., et al., *Applying cascaded convolutional neural network design further enhances automatic scoring of arthritis disease activity on ultrasound images from rheumatoid arthritis patients*. Ann Rheum Dis, 2020. **79**(9): p. 1189-1193.

46. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, 2017. **542**(7639): p. 115-118.
47. Deimel, T., D. Aletaha, and G. Langs, *Autoscra: Deep Learning to Automate Scoring of Radiographic Progression in Rheumatoid Arthritis*. Annals of the Rheumatic Diseases, 2020. **79**(Suppl 1): p. 39-40.
48. Esteva, A., et al., *Deep learning-enabled medical computer vision*. NPJ Digit Med, 2021. **4**(1): p. 5.
49. Fukae, J., et al., *Convolutional neural network for classification of two-dimensional array images generated from clinical information may support diagnosis of rheumatoid arthritis*. Sci Rep, 2020. **10**(1): p. 5648.
50. Khanna, N.N., et al., *Rheumatoid Arthritis: Atherosclerosis Imaging and Cardiovascular Risk Assessment Using Machine and Deep Learning-Based Tissue Characterization*. Curr Atheroscler Rep, 2019. **21**(2): p. 7.
51. Ureten, K., H. Erbay, and H.H. Maras, *Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network*. Clin Rheumatol, 2020. **39**(4): p. 969-974.
52. Lecun, Y. and Y. Bengio, *Convolutional Networks for Images, Speech, and Time-Series*. 1995.
53. Hubel, D.H. and T.N. Wiesel, *Receptive fields and functional architecture of monkey striate cortex*. J Physiol, 1968. **195**(1): p. 215-43.
54. Fukushima, K., *Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. Biol Cybern, 1980. **36**(4): p. 193-202.
55. Zeiler, M.D. and R. Fergus, *Visualizing and Understanding Convolutional Networks*. 2014, Springer International Publishing. p. 818-833.
56. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology*. Insights Imaging, 2018. **9**(4): p. 611-629.
57. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology*. Insights into Imaging, 2018. **9**(4): p. 611-629.
58. Le Cun, Y., et al. *Handwritten zip code recognition with multilayer networks*. IEEE Comput. Soc. Press.
59. Lee, H., et al., *Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks*. Commun. ACM, 2011. **54**: p. 95-103.
60. Zafar, A., et al., *A Comparison of Pooling Methods for Convolutional Neural Networks*. Applied Sciences, 2022. **12**(17): p. 8643.
61. Basha, S.H.S., et al., *Impact of fully connected layers on performance of convolutional neural networks for image classification*. Neurocomputing, 2020. **378**: p. 112-119.
62. Alzubaidi, L., et al., *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*. Journal of Big Data, 2021. **8**(1): p. 53.
63. Verma, R., V. Nagar, and S. Mahapatra, *Introduction to Supervised Learning*, in *Data Analytics in Bioinformatics*. 2021. p. 1-34.
64. Reed, R.D. and R.J. Marks, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. 1998: MIT Press.
65. Afaq, S. and S. Rao, *Significance Of Epochs On Training A Neural Network*. International Journal of Scientific & Technology Research, 2020. **9**: p. 485-488.
66. Sutton, R.S. and A.G. Barto, *Reinforcement learning: An Introduction*. Reinforcement learning: An introduction, 2nd ed. 2018, Cambridge, MA, US: The MIT Press. xxii, 526-xxii, 526.
67. van Otterlo, M. and M. Wiering, *Reinforcement Learning and Markov Decision Processes*, in *Reinforcement Learning: State-of-the-Art*, M. Wiering and M. van Otterlo, Editors. 2012, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 3-42.
68. Harmon, M.E. and S.S. Harmon, *Reinforcement Learning: A Tutorial*. 1997.
69. Li, Y., *Deep Reinforcement Learning*. arXiv pre-print server, 2018.

70. Ying, X., *An Overview of Overfitting and its Solutions*. Journal of Physics: Conference Series, 2019. **1168**: p. 022022.
71. Jabbar, H. and R.Z. Khan, *Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)*. Computer Science, Communication and Instrumentation Devices, 2015. **70**: p. 163-172.
72. Dietterich, T., *Overfitting and undercomputing in machine learning*. ACM computing surveys (CSUR), 1995. **27**(3): p. 326-327.
73. Wang, H., et al., *A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution, and Region Proposal*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020. **PP**: p. 1-18.
74. Lin, T.-Y., et al., *Feature Pyramid Networks for Object Detection*. 2017.
75. Lin, T.-Y., et al., *Focal Loss for Dense Object Detection*. 2018.
76. Zhou, Z., et al., *Image Processing: Facilitating Retinanet for Detecting Small Objects*. Journal of Physics: Conference Series, 2021. **1815**(1): p. 012016.
77. He, K., et al., *Deep Residual Learning for Image Recognition*. 2016. 770-778.
78. Barhamain, A.S., et al., *The journey of rheumatoid arthritis patients: a review of reported lag times from the onset of symptoms*. Open Access Rheumatol, 2017. **9**: p. 139-150.
79. Larsen, A., K. Dale, and M. Eek, *Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films*. Acta Radiol Diagn (Stockh), 1977. **18**(4): p. 481-91.
80. van der Heijde, D., et al., *Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis*. Rheumatology (Oxford), 1999. **38**(10): p. 941-7.
81. Sharp, J.T., et al., *Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities*. Arthritis Rheum, 1971. **14**(6): p. 706-20.
82. Sharp, J.T., et al., *How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis?* Arthritis Rheum, 1985. **28**(12): p. 1326-35.
83. van der Heijde, D., *How to read radiographs according to the Sharp/van der Heijde method*. J Rheumatol, 2000. **27**(1): p. 261-3.
84. Jung, A., *Imgaug*. 2020.
85. Yushkevich, P.A., et al., *ITK-SNAP*. 2021.
86. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. 2013.
87. Emmert-Streib, F., et al., *An Introductory Review of Deep Learning for Prediction Models With Big Data*. Frontiers in Artificial Intelligence, 2020. **3**: p. 4.
88. Duryea, J., S. Zaim, and F. Wolfe, *Neural network based automated algorithm to identify joint locations on hand/wrist radiographs for arthritis assessment*. Med Phys, 2002. **29**(3): p. 403-11.
89. Murakami, S., et al., *Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network*. Multimedia Tools and Applications, 2018. **77**(9): p. 10921-10937.
90. Welsing, P.M., et al., *The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis*. Arthritis Rheum, 2001. **44**(9): p. 2009-17.
91. Hirano, T., et al., *Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis*. Rheumatology Advances in Practice, 2019. **3**(2).
92. Wang, Z., et al., *An Efficient CNN for Hand X-Ray Overall Scoring of Rheumatoid Arthritis*. Complexity, 2022. **2022**: p. 5485606.