

Diplomarbeit

**MRT-BASIERTE
KNOCHENWACHSTUMSSCHÄTZUNG ZUR
PLANUNG VON EINGRIFFEN IN DER
KINDERORTHOPÄDIE – EINE RETROSPEKTIVE
PILOTSTUDIE**

eingereicht von

Judith Hörl

zur Erlangung des akademischen Grades

Doktor(in) der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt an der

Universitätsklinik für Kinder- und Jugendchirurgie,

Klinische Abteilung für Kinderorthopädie

unter der Anleitung von

Dr. med. univ. Michael Peter Novak

Priv.-Doz. Dr. med. univ. Tanja Kraus

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 24.06.2018

Judith Hörl eh

Danksagungen

An dieser Stelle möchte ich mich herzlich bei allen Personen bedanken, die mir bei der Erstellung meiner Diplomarbeit unterstützend zur Seite standen.

Zu Beginn möchte ich meinem Diplomarbeitsbetreuer Dr. Michael Novak danken, der die Arbeit vom Anfang bis zur Fertigstellung begleitet hat. Danke für die vielen hilfreichen Ratschläge, die Unterstützung und Motivation! Auch Priv.-Doz. Dr. Tanja Kraus möchte ich meinen Dank für die letzte Überprüfung aussprechen. Danke auch an Dr. Sebastian Tschauner für die Hilfe mit PACS.

Vielen Dank an das LBI CFI für die gute Zusammenarbeit, besonders an Dipl.-Ing. Dr. Martin Urschler für die Bereitstellung von Unterlagen und die vielen Erklärungen rund um die technischen Aspekte der Arbeit. Ich bin sehr dankbar für die interessanten Einblicke, die ich so bekommen konnte.

Herzlichen Dank an meine Familie und Freunde, die mich während des Studiums immer unterstützt und mir dieses erst ermöglicht haben. Mein besonderer Dank gilt Verena Hauck für die ausführlichen Korrekturen und Verbesserungsvorschläge sowie Gerald Krobath, der mir mit einem selbst programmierten Microsoft-Excel-Makro die Datenerfassung erleichtert hat.

Zusammenfassung

Einleitung: Das Skeletalter (SA) ist ein wichtiger Indikator für die Skelettreife, der in Orthopädie, Endokrinologie und Forensik Anwendung findet. Häufig verwendete Methoden zur SA-Bestimmung sind beispielsweise die Handskeletalterbestimmung nach Greulich und Pyle und die TW3-Methode, die allerdings auf Studienpopulationen des letzten Jahrhunderts basieren und darüber hinaus weitere Limitationen aufweisen. Eine softwarebasierte SA-Bestimmung anhand MRT-Aufnahmen des Kniegelenks hingegen verspricht objektive Ergebnisse bei gleichzeitig eingehaltener Strahlenhygiene und geringem Zeitaufwand. Auf die Wachstumsfugen des Kniegelenks kann ein überwiegender Anteil des Längenwachstums zurückgeführt werden, weshalb sich die Lokalisation sowohl zur SA-Bestimmung als auch zur darauf basierenden Voraussage der Erwachsenengröße eignet. In der vorliegenden Studie wurde die Hypothese geprüft, dass die softwarebasierte SA-Bestimmung anhand MRT-Aufnahmen des Kniegelenks möglich ist.

Methoden: Es erfolgten die Erstellung einer Datensammlung aus MRT-Aufnahmen des Kniegelenks sowie die Kalibrierung und Validierung einer selbstlernenden Software aus dem Bereich der Deep Learning Neural Networks. Bei der Datenakquirierung wurden Patientendaten und MRT-Aufnahmen aller Patientinnen/Patienten unter 18 Jahren gesammelt, die im Zeitraum 2000-2016 an der Abteilung für Radiologie der Medizinischen Universität Graz zwei oder mehr MRT-Untersuchungen des ipsilateralen Kniegelenks erhielten. Ausschlusskriterien umfassten schwere Grunderkrankungen und bestimmte kniebezogene Diagnosen mit schweren ossären strukturellen Veränderungen. Die SA-Bestimmung erfolgte durch Neural Networks mit unterschiedlichen Eigenschaften. Es wurde auch ein Overfitting-Experiment durchgeführt. Für die Experimente wurden MRT-Aufnahmen in T2-gewichteter Sequenz verwendet.

Ergebnisse: Es konnten Datensätze mit 317 MRT-Aufnahmen von 121 Patientinnen/Patienten gesammelt werden. Anhand der MRT-Aufnahmen von 83 Patientinnen/Patienten im Alter von 12 bis 18 Jahren konnte die SA-Bestimmung durchgeführt werden. Mit einem modifizierten Residual Network ergab sich ein stabiler Abfall des mittleren absoluten Fehlers (MAE). Der niedrigste MAE, der in der Validierung erzielt werden konnte, betrug 1,04 Jahre.

Diskussion: Die Ergebnisse sprechen für einen Lernfortschritt der Software. Das erfolgreiche Overfitting-Experiment zeigte, dass die zur SA-Bestimmung notwendige Information in den MRT-Aufnahmen enthalten ist. Somit konnte die Hypothese, dass die SA-Bestimmung auf diese Weise möglich ist, bestätigt werden. Die Durchführung weiterer Studien mit größeren Datenmengen zur Steigerung der Genauigkeit und Aussagekraft der Methode erscheint daher sehr empfehlenswert. In Folgestudien könnte auch die Voraussage der Erwachsenengröße durch die Software erprobt werden.

Abstract

Introduction: Bone age (BA) is an indicator of skeletal maturity used in orthopaedics, endocrinology and forensics. Commonly used methods for its determination are the radiological BA assessment of hand and wrist by Greulich and Pyle and the TW3 method, which are based on outdated study populations and are otherwise disadvantageous. A software-based assessment of BA utilising MRI images of the knee joint however promises objective results with a small expenditure of time while remaining radiation-free. A predominant proportion of increase in body length can be attributed to the growth plates of the knee. Therefore, this region seems suitable for both BA assessment and prediction of adult height. The present study investigates the hypothesis that software-based BA determination utilising MRI images of the knee is possible.

Methods: A data collection with MRI images was created and the calibration and validation of a deep learning neural network software was conducted. During data acquisition, patient data and MRI scans of patients under the age of 18 years who received two or more MRI scans of the ipsilateral knee joint were collected. Exclusion criteria included severe underlying diseases and certain knee-related diagnoses with extensive structural changes of bone morphology. The BA assessment was carried out by neural networks with different properties. An overfitting experiment was also performed. For the experiments, MRI images in T2 -weighted sequence were used.

Results: Data sets of 317 MRI images from 121 patients were collected. Utilising the MRI images of 83 patients aged 12 to 18 years, the BA determination was performed. The validation of a modified residual network resulted in a stable decrease in mean absolute error (MAE). The lowest MAE was 1.04 years.

Conclusion: The results are in favour of the software's learning progress. The successful overfitting experiment indicates that the MRI images contain information necessary for this task. Thus, the hypothesis that BA determination is possible in this way has been confirmed. Nonetheless, further studies with larger amounts of data are necessary to increase the accuracy and validity of the software. In subsequent studies, the prediction of adult height could also be tested.

Inhaltsverzeichnis

Danksagungen	ii
Zusammenfassung	iii
Abstract	iv
Inhaltsverzeichnis	v
Glossar und Abkürzungen.....	vii
Abbildungsverzeichnis	viii
1 Einleitung.....	10
1.1 Skeletalter und Entwicklung.....	10
1.1.1 Skeletalter und chronologisches Alter	10
1.1.2 Skeletalter und Pubertät	11
1.1.3 Einflüsse auf das Skeletalter	12
1.1.4 Anwendungsbereiche	12
1.2 Bestimmung des Skeletalters	14
1.2.1 Handskeletalterbestimmung.....	14
1.2.1.1 Methode nach Greulich und Pyle.....	14
1.2.1.1.1 Hintergründe.....	15
1.2.1.1.2 Probleme und Limitationen	15
1.2.1.2 Methode nach Tanner Whitehouse	18
1.2.1.2.1 Entstehung	18
1.2.1.2.2 Hintergründe und Studienpopulation	19
1.2.1.2.3 TW3 im Vergleich	21
1.2.1.3 Die FELS-Methode	22
1.2.2 Andere Röntgen-basierte Methoden der Altersbestimmung.....	23
1.2.2.1 Dental Age Assessment	23
1.2.2.2 Risser-Zeichen	23
1.2.2.3 Skeletalterbestimmung am Knie.....	24
1.2.3 MRT-basierte Methoden	25
1.2.4 Automatisierte Methoden.....	27
1.3 Voraussage der Erwachsenengröße	29
1.3.1 Anwendungsbereiche	29
1.3.2 Methoden zur Voraussage der Erwachsenengröße	32
1.3.2.1 Voraussage der Erwachsenenstehhöhe nach Bayley und Pinneau	32
1.3.2.2 Die Roche-Wainer-Thissen-Methode.....	33
1.3.2.3 Verbleibendes Wachstum der unteren Extremitäten nach Anderson ..	34
1.3.2.4 Moseley Straight-Line Graph.....	35
1.3.2.5 Multiplier Method	36
1.4 Grundlagen des maschinellen Lernens und Deep Learning	39
1.4.1 Erklärung der Arbeitsweise	39
1.4.2 Aufbau eines Neural Network	40
1.4.3 Training und Validierung	40
1.4.4 Probleme und Lösungsstrategien	41
1.5 Hypothese, Fragestellung und Zielsetzung.....	43
2 Material und Methoden	45
2.1 Studiendesign und -ablauf.....	45
2.1.1 Zielgrößen	46
2.2 Datenakquirierung.....	47

2.2.1	Ein- und Ausschlusskriterien.....	47
2.2.2	Datensammlung	49
2.3	Altersschätzung durch Neural Network	51
3	Ergebnisse – Resultate.....	52
3.1	Ergebnisse der Datenakquirierung.....	52
3.2	Registrierung und Auswahl der Daten.....	55
3.3	Ergebnisse der Altersschätzung.....	58
4	Diskussion	63
4.1	Limitationen bestehender Methoden und neue Ansätze	63
4.1.1	Methoden zur Skelettalterbestimmung.....	63
4.1.2	Methoden zur Voraussage der Erwachsenengröße	64
4.1.3	Neue Ansätze und vergleichbare Studien	64
4.2	Limitationen und Stärken der Studie.....	67
4.2.1	Ergebnisse der Pilotstudie	67
4.2.2	Methodik und Studiendesign	70
4.2.2.1	Datenakquirierung und Studienpopulation	70
4.2.2.2	Schätzung des SA.....	73
4.3	Erkenntnisse und Schlussfolgerungen.....	76
4.4	Ausblick.....	78
5	Literaturverzeichnis	79

Glossar und Abkürzungen

MRT – Magnetresonanztomographie

BA – bone age

CA – chronologisches Alter

SA – Skeletalter

TW1 – Tanner Whitehouse 1

TW2 – Tanner Whitehouse 2

TW3 – Tanner Whitehouse 3

SMS – Skeletal Maturity Score

RUS-Score – Radius, Ulna and Short bone score

CASAS – Computer-Assisted Skeletal Age Scoring System

RMSD – root mean square deviation, Wurzel der mittleren quadratischen Abweichung

CNN – Convolutional Neural Network

PMH – per cent of mature height-Wert

RWT-Methode – Roche-Wainer-Thissen-Methode

M-SLG – Moseley Straight-Line Graph

KI – künstliche Intelligenz

AI – Artificial Intelligence

ROI – Region of Interest

CT – Computertomographie

LBI CFI – Ludwig Boltzmann Institut für Klinisch-Forensische Bildgebung

EDV – elektronische Datenverarbeitung

PACS – Picture Archiving and Communication System

SAP – System Applications and Products in Data Processing

DICOM – Digital Imaging and Communications in Medicine

LKH – Landeskrankenhaus

ResNet – Residual Network

MAE – mean absolute error, mittlerer absoluter Fehler

mAP – mean average precision, mittlere durchschnittliche Präzision

Abbildungsverzeichnis

Abbildung 1: Knochen der beiden Scores: Für den RUS-Score werden Radius, Ulna sowie Metakarpalknochen und Phalangen des ersten, dritten und fünften Fingers beurteilt. Der Karpalscore setzt sich aus den Scores der Handwurzelknochen (exklusive Os pisiforme) zusammen (entnommen aus (38) mit Erlaubnis von Elsevier).....	20
Abbildung 2: Korrelation zwischen CA und mittels MRT des Kniegelenks bestimmtem SA (entnommen aus (60) mit Erlaubnis von Wolters Kluwer Health, Inc.).	25
Abbildung 3: Vergleich der Bildmodalitäten: Verschiedene Stadien der Entwicklung des Kniegelenks in MRT-Aufnahmen (oben) und auf Röntgenbildern (unten)(entnommen aus (59) mit Erlaubnis von Elsevier).	26
Abbildung 4: Orthopädische Korrekturplanung: Zur Planung der wachstumslenkenden Achskorrektur bei Genua valga mittels temporärer Hemiepiphyseodese (Eight-Plates) ist die Kenntnis des verbleibenden Wachstumspotentials unerlässlich (von der Abteilung für Kinderorthopädie der Universitätsklinik für Kinder- und Jugendchirurgie zur Verfügung gestellt).	30
Abbildung 5: Moseley Straight-Line Graph (entnommen aus (84) mit Erlaubnis von Wolters Kluwer Health, Inc.).	36
Abbildung 6: Tabellen altersspezifischer Multiplikatoren: Femorale Multiplikatoren für Knaben (links) und tibiale Multiplikatoren für Mädchen (rechts)(entnommen aus (85) mit Erlaubnis von Wolters Kluwer Health, Inc.).....	37
Abbildung 7: Schema eines klassischen Neural Network: An den Knoten werden Entscheidungen getroffen, die an die folgenden Ebenen weitergegeben werden. Die Knoten einer Schicht sind mit allen Knoten der nächsten Schicht verbunden. Die Gewichte entlang dieser Verbindungen sind die Parameter, die im Training der Software adaptiert werden (eigene Darstellung).	40
Abbildung 8: Schematische Darstellung unterschiedlicher Neural Networks: Plain Network (links) und Residual Network (rechts). Das Residual Network weist zusätzliche Verbindungen zwischen nicht benachbarten Ebenen auf (eigene Darstellung).....	42
Abbildung 9: Zusammensetzung der Studienpopulation: Unter 121 Patientinnen/Patienten waren 59 Mädchen und 61 Knaben. Von 317 MRT-Aufnahmen zeigten 146 das linke und 171 das rechte Knie (eigene Darstellung).	52
Abbildung 10: Altersverteilung der Studienpopulation: Von den 317 MRT-Untersuchungen wurde ein großer Anteil im Alter von 12 bis 16 Jahren durchgeführt. Besonders jüngere Lebensalter sind unterrepräsentiert. Das durchschnittliche Alter bei der Untersuchung lag bei 14,24 Jahren (eigene Darstellung).....	53
Abbildung 11: Anzahl der MRT-Untersuchungen pro Knie: Da bei mehreren Patientinnen/Patienten beide Kniegelenke untersucht wurden, werden an dieser Stelle die Untersuchungen pro Knie angegeben. In den meisten Fällen lagen nur zwei MRT-Aufnahmen vor. Nur in vier Fällen waren fünf oder mehr Untersuchungen durchgeführt worden. Für die SA-Bestimmung wurde allerdings nur eine MRT-Aufnahme pro Patientin/Patient verwendet (eigene Darstellung).	53
Abbildung 12: Altersverteilung der Studienpopulation nach Registrierung (von LBI CFI zur Verfügung gestellt).	55
Abbildung 13: Altersverteilung der Studienpopulation nach Registrierung und Ausschluss dreier Patientinnen/Patienten (von LBI CFI zur Verfügung gestellt).....	56
Abbildung 14: Altersverteilung von Trainings- und Validierungsset (von LBI CFI zur Verfügung gestellt).	56

Abb. 15: Beispiel einer MRT-Aufnahme nach erfolgter Registrierung mit rot eingezeichneter ROI bzw. Volume of Interest (von LBI CFI zur Verfügung gestellt).	57
Abbildung 16: Veränderung des mittleren absoluten Fehlers (MAE) in Experiment 1 in Training (blau) und Validierung (orange). Es zeigt sich keine signifikante Verbesserung in der Validierung (von LBI CFI zur Verfügung gestellt).....	58
Abbildung 17: Veränderung des MAE in Experiment 4 (kontrolliertes Overfitting-Experiment) in Training (blau) und Validierung (orange)(von LBI CFI zur Verfügung gestellt).	59
Abbildung 18: Veränderung des MAE in Experiment 5 in Training (blau) und Validierung (orange). Es zeigte sich ein stabiler Abfall des MAE (von LBI CFI zur Verfügung gestellt)..	60
Abbildung 19: Veränderung des MAE in Experiment 6 in Training (blau) und Validierung (orange). Es zeigte sich ein Ausreißer um den 20. Durchlauf, der niedrigste MAE wurde kurz davor erreicht (von LBI CFI zur Verfügung gestellt).....	61
Abbildung 20: Veränderung des MAE in Experiment 9 in Training (blau) und Validierung (orange). Es zeigten sich Ausreißer um 15. sowie zwischen 30. und 40. Durchlauf, niedrigster MAE erst um den 40. Durchlauf erreicht (von LBI CFI zur Verfügung gestellt).	61
Abbildung 21: Veränderung des MAE in den Experimenten 8 (links) und 5 (rechts) in Training (blau) und Validierung (orange)(von LBI CFI zur Verfügung gestellt).	62

1 Einleitung

1.1 *Skeletalter und Entwicklung*

1.1.1 **Skeletalter und chronologisches Alter**

In der Medizin unterscheidet man zwischen dem chronologischen Alter (CA) eines Menschen, das durch das Geburtsdatum bestimmt wird, und dem Skeletalter (SA), welches ein wichtiger Indikator für die Skelettreife ist (1). Normalerweise stimmen diese beiden Werte überein, jedoch kann es durch unterschiedliche Ursachen zu Abweichungen des Skeletalters vom chronologischen Alter kommen. Dieser Umstand spielt sowohl in der pädiatrischen Orthopädie als auch in Endokrinologie und Forensik eine wichtige Rolle. Daher wurden im Laufe der Zeit verschiedene Methoden zur Bestimmung des Skeletalters entwickelt (1).

Um Gedeihstörungen bei pädiatrischen Patientinnen/Patienten möglichst früh zu erkennen, ist es notwendig, bestimmte Werte, die sich im Laufe der Entwicklung verändern, zu messen und mit den Werten anderer Kinder zu vergleichen. Als Beispiele solcher Werte sind Körpergröße und Körpergewicht zu nennen, die in der Praxis auf sogenannten Perzentilenkurven eingetragen werden (1). So wird ein Vergleich der Parameter der Patientin/des Patienten mit Werten der Normalverteilung gesunder Kinder im gleichen Alter ermöglicht und Abweichungen können erkannt werden. Allerdings sind Größe und Gewicht zwar Hinweise, allein aber bei Weitem nicht ausreichend, um Störungen in der Entwicklung eines Kindes sicher aufzudecken, wie schon Greulich und Pyle erkannten (2). Sie begründeten dies damit, dass der Eintritt in die Pubertät und der damit verbundene Wachstumsschub bei Kindern nicht zwingend im gleichen Alter erfolgen müssen. Demnach können Kinder zwar gleich alt, aber unterschiedlich groß und schwer sein – das Alter ist also nicht sehr aussagekräftig. Da das Größenwachstum bei Kindern, die frühzeitig in die Pubertät kommen, hormonell bedingt auch entsprechend früher stagniert, erreichen diese Kinder ihre endgültige Größe früher (3, 4). Im Extremfall einer Pubertas praecox wäre das Kind in den Jahren nach seiner Pubertät zwar kleiner als Gleichaltrige, die sich zu dem Zeitpunkt noch im Wachstum befinden, aber hinsichtlich SA und körperlicher Entwicklung schon weiter fortgeschritten (2). Bei einer verzögert eintretenden oder ausbleibenden Pubertät gilt das Gegenteil: Die Erwachsenengröße wird

später erreicht, da bis zum Ende der Pubertät mehr Zeit für das Längenwachstum verfügbar ist. An diesen Beispielen wird deutlich, dass Körpergröße und Körpergewicht nicht ausreichen, um die körperliche Entwicklung von Kindern zu beschreiben. Andere individuelle Faktoren wie beispielsweise die Körpergröße der Eltern haben ebenfalls einen Einfluss auf das Wachstum eines Kindes (1). Weiters weisen Größen- und Gewichtsdatensammlungen Limitationen auf, wenn das Kind, dessen Werte man anhand der Daten vergleichen will, aus einer anderen Bevölkerungsgruppe stammt als die Population der Datensammlung. Ein verlässliches und von derartigen Variablen unabhängiges Maß für die körperliche Entwicklung hinsichtlich des Längenwachstums während der Wachstumsphase eines Kindes ist daher wünschenswert.

1.1.2 Skeletalter und Pubertät

Das Skeletalter kann nicht nur wertvolle Informationen zum Reifegrad des kindlichen Skelettes liefern, sondern erlaubt aufgrund der engen Zusammenhänge zwischen dem Knochenwachstum und der Reifung anderer Organsysteme auch Rückschlüsse auf den Entwicklungsstatus des gesamten Körpers (2, 4). Besonders das endokrine System ist diesbezüglich erwähnenswert, da mit der Pubertät ein Wachstumsschub und die Entwicklung sekundärer Geschlechtsmerkmale einhergehen (1). Dies führt zu signifikanten, äußerlich erkennbaren Veränderungen des Körpers. Schon Simmons und Greulich konnten eine starke Korrelation zwischen Skelettreife und Pubertät beobachten, als sie feststellten, dass Mädchen im Jahr vor ihrer Menarche die stärkste Zunahme der Körpergröße aufweisen und diese mit dem SA besser voraussagbar ist als mit Körpergröße oder -gewicht (3). Man spricht von einem präadoleszenten Wachstumsschub, der bei Knaben im Vergleich zu Mädchen später auftritt. Diesem folgt ein Abfall der Wachstumsgeschwindigkeit, bis das Größenwachstum letztendlich mit dem Epiphysenschluss stagniert (2). Somit kann mittels Bestimmung des SA beispielsweise eine Schätzung des Zeitpunktes dieses Wachstumsschubes oder auch der Wahrscheinlichkeit, in einem gewissen Zeitraum die Skelettreife bzw. Erwachsenengröße zu erreichen, erfolgen (5). Das SA kann natürlich nicht nur in dem kurzen Zeitraum vor und während der Pubertät bestimmt werden, sondern von Geburt an bis zum Abschluss des Wachstums (2).

1.1.3 Einflüsse auf das Skeletalter

Zahlreiche Faktoren können bewirken, dass das SA vom CA oder von dem SA anderer Kinder abweicht. Mädchen haben generell ein höheres SA als Knaben, der Epiphysenschluss erfolgt ungefähr zwei Jahre früher (4). Das Geschlecht spielt also eine nicht unerhebliche Rolle, ebenso wie genetische Faktoren. Allerdings haben Umweltfaktoren einen nicht weniger großen Einfluss auf Wachstum und Knochenalter, wie eine Studie von Greulich impliziert (6). Dabei wurden in den USA geborene Kinder japanischer Herkunft mit kaukasischen Kindern aus Cleveland verglichen. Die beiden Gruppen waren sich viel ähnlicher als dieselbe Gruppe von Kindern aus Cleveland und eine in Japan lebende Population, die im Vergleich ein dezeleriertes SA aufwies (6-8). Die unterschiedlichen Umweltbedingungen hatten in diesen Studien also einen stärkeren Einfluss als die ethnische Herkunft (7). Das SA beeinflussende Umweltfaktoren können Strahlung, Infektionen und nicht zuletzt die Ernährung sein (2, 4, 9). Populationen mit hohem sozioökonomischem Status weisen tendenziell höhere Körpergröße und frühere Reife auf (9). Niedriger sozialer Status und mangelhafte Ernährung zeigten eine verzögerte Skelettreife (5).

Nicht nur Infektionen, sondern auch andere Erkrankungen können das SA im Vergleich zum chronologischen Alter verzögern oder vorantreiben. Dazu gehören endokrine Störungen, genetische Syndrome (z.B. Trisomien, Beckwith-Wiedemann-Syndrom, Sotos-Syndrom), Reifungsstörungen, Adipositas, Rachitis und andere (4). Schwere Krankheiten oder andere Ursachen können zu temporären Wachstumsunterbrechungen führen, die sogar Spuren an den Knochen hinterlassen. Greulich und Pyle bezeichneten diese am Röntgenbild sichtbaren Linien als „Scars of interrupted growth“ (2). Diese erlauben auch die ungefähre zeitliche Zuordnung des Auftretens der Störung.

1.1.4 Anwendungsbereiche

Das SA wird häufig und für viele unterschiedliche Fragestellungen erhoben.

Kinder, die durch Wachstumsstörungen hinsichtlich unter- oder überdurchschnittlicher Körpergröße auffallen, werden meist endokrinologisch abgeklärt (4). Zu dieser Abklärung gehört auch eine SA-Bestimmung. Die häufigsten Fragestellungen sind verzögert oder frühzeitig eintretende Pubertät, Klein- und Großwuchs, adrenogenitales Syndrom, Hypo-

und Hyperthyreose und Somatotropinmangel (4, 10). Eine frühe Diagnose ist wichtig, um rechtzeitig therapeutisch eingreifen zu können. Ob und in welchem Alter man eine Therapie beginnt, hängt oftmals vom SA ab. Im Falle von Therapien, die das Wachstum regulieren sollen, ist eine Überwachung auch mittels SA angebracht (11). So kann der Therapieerfolg überprüft und verbessert werden.

In der Orthopädie wird eine Bestimmung des SA oft in Kombination mit Voraussage der Erwachsenengröße durchgeführt. Vor korrigierenden chirurgischen Eingriffen von beispielsweise Skoliose oder Beinlängendifferenzen ist es zur Planung notwendig festzustellen, ob das Wachstum bereits abgeschlossen ist oder ob noch Veränderungen zu erwarten sind (4, 12). Manche Eingriffe können erst bei abgeschlossenem Wachstum durchgeführt werden, manche nur solange das Kind noch wächst (13). Das SA entscheidet bei bestimmten Achsdeformitäten mit darüber, ob eine Osteotomie oder ein schonenderes wachstumslenkendes Verfahren wie die temporäre Hemiepiphyseodese durchgeführt wird (14).

Auch in der forensischen Anthropologie hat die SA-Bestimmung durch Migrationsbewegungen in den letzten Jahren an Bedeutung gewonnen (15). Wenn das Alter eines jugendlichen Flüchtlings unklar ist, kann es aufgrund rechtlicher Gegebenheiten notwendig werden, zu beweisen oder zu widerlegen, dass sie oder er minderjährig ist. Dazu wird unter anderem das SA herangezogen. So auch im Falle von Adoptionen oder der Frage nach der Mündigkeit jugendlicher Rechtsbrecherinnen/Rechtsbrecher unbekanntes Alters (16). Ergebnisse solcher Untersuchungen haben auch nichtmedizinische Konsequenzen (15). Untersuchungen des SA werden nicht nur an lebenden Menschen, sondern auch bei archäologischen Leichenfunden durchgeführt.

Weiters ist noch erwähnenswert, dass das Knochenwachstum auch in der Odontologie bei Patientinnen/Patienten, die Zahnimplantate benötigen, eine Rolle spielt (5). All diese Bereiche würden von einer Verbesserung der Messung des SA profitieren.

1.2 Bestimmung des Skeletalters

An dieser Stelle sollen die gängigsten und am weitesten verbreiteten Methoden zur Bestimmung des SA erläutert werden. Aber auch jüngere Verfahren, die weniger etabliert, jedoch für diese Studie von Bedeutung sind, werden im Folgenden beschrieben.

1.2.1 Handskeletalterbestimmung

Die bekannteste und am häufigsten verwendete Methode zur Bestimmung des SA ist die Handskeletalterbestimmung nach Greulich und Pyle, deren Entwicklung bereits 1929 begann (2, 17). Davon abgesehen existieren auch andere Methoden zur Bestimmung des Handskeletalters (1, 18, 19). Dabei handelt es sich um Atlasvergleichsmethoden oder Scores, die Röntgenbilder der linken Hand verwenden. Das bedeutet, dass von der zu untersuchenden Patientin/dem zu untersuchenden Patienten ein Handröntgen der linken Hand in antero-posteriorem Strahlengang angefertigt wird. Dieses wird von der/dem Beurteilenden, die/der in der Regel Radiologin/Radiologe, Orthopädin/Orthopäde oder Pädiaterin/Pädiater ist, mit den Bildern oder Scores eines bereits bestehenden Atlas verglichen. In dem Atlas nach Greulich und Pyle finden sich Röntgenbilder linker Hände von Kindern und Jugendlichen ansteigenden Alters, mit deren Hilfe eine Schätzung des Skeletalters der Patientin/des Patienten ermöglicht wird (2).

1.2.1.1 Methode nach Greulich und Pyle

Der „Radiographic Atlas of Skeletal Development of the Hand and Wrist“ von Greulich und Pyle entstand aus einem Projekt, dessen Planung von 1929 bis 1931 dauerte (2). Kinder im Alter von 3 Monaten bis 14 Jahren wurden in die Studie eingeschlossen und in einem longitudinalen Programm bis 1942 begleitet. Im ersten Lebensjahr wurden alle drei Monate Röntgenbilder der linken Extremitäten angefertigt (Schulter, Ellbogen, Hand, Hüfte, Knie und Fuß). Bis zum Alter von fünf Jahren erfolgten diese Aufnahmen halbjährlich, danach jährlich (2).

Doch wie setzte sich die Studienpopulation zusammen? Die Probandinnen/Probanden wurden nur auf Antrag ihrer Kinderärztinnen/Kinderärzte für die Studie rekrutiert, weshalb die Autoren des Atlas vermuteten, dass die meisten aus höheren sozialen Schichten stammten (2). Alle Probandinnen/Probanden waren Kaukasierinnen/Kaukasier

mit nordeuropäischer Abstammung, die aber in den Vereinigten Staaten von Amerika geboren wurden. Ausschlusskriterien waren Krankheiten nicht nur physischer, sondern auch psychischer Art. Um diese zu überwachen, wurden die Krankengeschichten überprüft und zu jedem Messzeitpunkt Körpergröße und Gewicht ermittelt. Schlussendlich konnten 2 bis 21 Handröntgenaufnahmen von 1000 Probandinnen/Probanden gesammelt werden. Die für ihr Alter repräsentativsten Röntgenbilder wurden ausgewählt und als „Standards“ in den Atlas aufgenommen. Aufgrund der raschen morphologischen Veränderung im Rahmen der Pubertät wurde ein zusätzlicher Standard im Alter von 13,5 Jahren bei Mädchen und im Alter von 15,5 Jahren bei Knaben eingefügt (2).

1.2.1.1.1 Hintergründe

Damit eine Altersbestimmung mittels Vergleich von Röntgenaufnahmen möglich ist, muss es im Wachstum zu charakteristischen Veränderungen der Knochen kommen. Die einzelnen Ossifikationszentren verändern sich in Form und Größe und verknöchern in einer bestimmten Reihenfolge (2). Der Schluss der Epiphysenfuge markiert den Abschluss des Wachstums eines Knochens (1, 2). Auf diese Weise lassen sich Wachstumsstadien einteilen. Greulich und Pyle verwendeten für diese von der Reife abhängigen Schritte der Veränderung einzelner Knochen den Begriff „maturity indicators“ (2). Die Hand besteht aus vielen einzelnen Knochen, was für diese Methode günstig ist. Eine Korrelation zwischen den Ergebnissen eines Handröntgens und der Skelettreife des gesamten Skeletts ermöglicht Rückschlüsse auf die körperliche Entwicklung (20). Um die Strahlendosis zu reduzieren, wird nur eine Seite untersucht. Da die meisten Menschen Rechtshänder sind und Empfehlungen vorlagen, die linke Seite für Messungen zu bevorzugen, wählten Greulich und Pyle die linke Hand aus (2).

1.2.1.1.2 Probleme und Limitationen

Obwohl seit der Entwicklung des Atlas von Greulich und Pyle schon viele Jahre vergangen sind, wird diese Methode zur Bestimmung des SA bis heute am häufigsten verwendet (1, 17, 21). Ein Röntgen der Hand ist in der Regel schnell verfügbar und laut einer Studie von King et al. auch in 1,4 Minuten ausgewertet (22). Die einzelnen Knochen der Hand entwickeln sich wie bereits erwähnt in einer bestimmten Reihenfolge. Diese kann jedoch zum Beispiel durch Krankheit gestört werden, sodass ein oder mehrere

Ossifikationszentren vom SA der gesamten Hand abweichen (2). Um diese Abweichungen auszugleichen, fordern Greulich und Pyle, dass stark abweichende Ossifikationszentren im Befund zusätzlich zum SA erwähnt werden. Denn würde man jedem einzelnen Knochen einen Wert zuweisen und diese Werte dann addieren oder den Mittelwert berechnen, könnte ein retardierter Knochen das Ergebnis verfälschen. Die Methode nach Greulich und Pyle umgeht dieses Problem (2). Dennoch wird eine korrekte Bestimmung des SA durch derartige Abweichungen erschwert.

1.2.1.1.2.1 Anwendung in anderen Populationen

Ein weiterer Punkt ist die Anwendbarkeit an anderen Populationen als der Studienpopulation. Heute erreichen Kinder früher die Pubertät als dies vor einigen Jahrzehnten der Fall war (23). Der Einfluss dieser veränderten Umweltbedingungen auf das SA wurde bereits oft untersucht, ebenso die Unterschiede zu Populationen anderer Herkunft (15, 24-32). Loder et al. fanden besonders bei Mädchen afroamerikanischen Ursprungs Abweichungen des SA vom CA und schlossen daraus, dass die Methode nicht auf alle Populationen anwendbar sei (25). Auch Ontell et al. stellten signifikante Abweichungen vom CA fest, auch bei Probandinnen/Probanden kaukasischer Herkunft. So war das SA präpubertärer Knaben im Vergleich zum CA verzögert (26). In einer französischen Population 10- bis 19-Jähriger wurde eine geringe Überschätzung des CA festgestellt, die allerdings nicht von statistischer Signifikanz war (15). Ebenso gering waren die Abweichungen in einer australischen Studie, was für die Anwendbarkeit der Methode in Australien spricht (27). In einer pakistanischen Population im Alter von 4,5 bis 9,5 Jahren wurde das CA durch das ermittelte SA signifikant unterschätzt (28). Die Variationen waren bei Knaben stärker ausgeprägt. Daher raten die Autoren des Artikels, die Methode nicht zu verwenden, wenn besondere Präzision erforderlich ist. Durch die starke Korrelation von SA und CA eignet sie sich allerdings gut für Verlaufskontrollen (28). Mora et al. untersuchten eine amerikanische Population und verglichen die Ergebnisse der Probandinnen/Probanden europäischer Abstammung mit denen der Probandinnen/Probanden afrikanischer Abstammung. Vor der Pubertät hatten die Kinder europäischer Abstammung ein im Vergleich verzögertes SA, bei den Knaben drehte sich dieser Effekt postpubertär um (29). In einer italienischen Studie wurde deutlich, dass die Methode zwar für klinische Fragestellungen nützlich, aber in der Forensik nicht sehr präzise ist. Gerade im Alter von 14 bis 18 Jahren, welches von praktischer Relevanz ist,

gab es signifikante Abweichungen (30). In einer anderen Studie fanden Santoro et al. in ihrer italienischen Studienpopulation allerdings Ergebnisse, die für die Verlässlichkeit der Methode nach Greulich und Pyle sprechen (31). In einer österreichischen Population wurden starke Abweichungen während und nach der Pubertät gefunden (32). Auch Gröll et al. fanden bei zentraleuropäischen Kindern eine Unterschätzung des CA mit Abweichungen des gemessenen SA vom wahren CA von bis zu einem Jahr (24). Diese Werte liegen jedoch innerhalb der von Greulich und Pyle beschriebenen Bandbreite der Variation des normalen SA, weshalb die Autoren die Methode als anwendbar erachteten (24). Tatsächlich wussten Greulich und Pyle bereits 1959, dass man nicht erwarten konnte, dass die von ihnen ausgesuchten Standardröntgenbilder zu jeder Population präzise passen würden. Das liegt nicht zuletzt an der Variabilität, die durch genetische Unterschiede zustande kommt (2). Wenn man also die Handskeletalterbestimmung nach Greulich und Pyle anwendet, sollte man sich auch ihrer Limitationen bewusst sein und eventuell sogar Anpassungen vornehmen. Ein Atlas mit aktuelleren, an die zu untersuchende Population angepassten Bildern würde diesem Problem Abhilfe schaffen (30). 1988 unternahmen Cole et al. den Versuch, die Handskeletalterbestimmung nach Greulich und Pyle für mitteleuropäische Kinder zu verbessern (33). Dabei wurden Handröntgenbilder von 97 Kindern verwendet, um die Abweichungen des CA von dem SA zu berechnen. Für jedes SA wurde das durchschnittliche mittlere CA der Probandinnen/Probanden zu diesem Zeitpunkt angegeben. Bei der Erstellung eines SA-Befundes können ergänzend zu dem SA nach Greulich und Pyle auch die Korrekturwerte nach Cole et al. angegeben werden (1, 33).

1.2.1.1.2.2 Subjektivität der Methode

Bei der Handskeletalterbestimmung nach Greulich und Pyle handelt es sich um eine subjektive Methode. Die Gesamteindrücke mehrerer Röntgenbilder werden von der Untersucherin/dem Untersucher miteinander verglichen, weshalb es naheliegend ist, dass es zu Ungenauigkeiten kommen kann (1, 2). Den Autoren zufolge ist zu erwarten, dass zwei Untersucherinnen/Untersucher mit unterschiedlicher Erfahrung trotz dieser subjektiven Komponente Schätzungen abgeben, die nicht mehr als sechs Monate auseinander liegen (2). Es wurden bereits einige Studien durchgeführt, die die Intra- und Interobserverabweichung überprüften. In der Studie von Tisè et al. wurden die Röntgenbilder von zwei Ärzten ausgewertet (30). Keine statistisch signifikanten Intra-

oder Interobserverabweichungen wurden gefunden, in einem großen Teil der Fälle stimmten die Ergebnisse überein. In 10,3% der Fälle lagen sie sechs Monate auseinander (30). Weitere Studien untermauern die Aussage, dass die Intra- und/oder Interobserverabweichung gering ist (24, 27, 34, 35). Außerdem zeigt sich auch ein Lerneffekt: Je erfahrener die Untersucherin/der Untersucher ist, desto geringer ist auch die Intraobserverabweichung (22, 24, 35). Des Weiteren ist die Methode für Unerfahrene einfach zu erlernen (2, 30, 34). In einer Umfrage von Breen et al. wurden Radiologinnen/Radiologen zum Thema Schätzung des SA befragt (17). Es zeigte sich, dass die meisten Befragten die Methode nach Greulich und Pyle wählten. Auch das Vertrauen in die verwendeten Methoden wurde anonym erfragt. Dieses Vertrauen hing in erster Linie vom CA der Patientinnen/Patienten ab. Bei jüngeren Kindern herrschte unter den Befragten die größere Unsicherheit: Von den 441 Teilnehmerinnen/Teilnehmern gaben 34% an, bei der SA-Schätzung von Säuglingen unsicher zu sein. Bei Kindern im Alter von ein bis drei Jahren waren es 22%, ab drei Jahren nur mehr 7%. Breen et al. führen das auf die nur geringen Veränderungen im Handröntgen in diesem Alter zurück (17). Eine Methode, die bei Säuglingen und Kleinkindern keine derartigen Limitationen aufweist, ist wünschenswert.

1.2.1.2 Methode nach Tanner Whitehouse

Eine weitere Methode zur Skeletalterbestimmung wurde von Tanner, Whitehouse und Healy entwickelt. Wie auch bei der Methode nach Greulich und Pyle werden Röntgenbilder der linken Hand verwendet. Das Verfahren wurde mehrmals überarbeitet und verbessert (18).

1.2.1.2.1 Entstehung

Die Entwicklung der ersten Version der Tanner Whitehouse Methode für die Schätzung des SA (TW1) wurde 1962 abgeschlossen (18). Der Neuheitswert bestand darin, dass die Methode auf einer mathematischen Basis beruhte und versuchte, die Reife quantitativ als Score und nicht direkt als SA anzugeben. 1983 erschien die erste Überarbeitung der Methode, TW2. Die zu den Stadien gehörigen Scores wurden verändert und Mädchen und Knaben getrennt bewertet. Die Studienpopulation blieb jedoch dieselbe. Die aktuellste Version, TW3, wurde 2001 veröffentlicht. Scores und Stadien blieben unverändert, die Referenzwerte aber wurden an aktuellere Beobachtungen der

Wachstumsgeschwindigkeit angepasst. Dazu wurden Daten aus Nordamerika und Europa verwendet. Die Population, auf der die Gleichungen zur Vorhersage der Erwachsenengröße basieren, besteht aus Kindern der First Zurich Longitudinal Growth Study, die auch die Größe der Eltern erfasste. Das mittels TW2 und TW3 erhobene SA unterscheidet sich bei jüngeren Kindern kaum, ab 10-11 Jahren finden sich Unterschiede von ca. einem Jahr, wobei TW3 ein höheres SA ergibt (18).

1.2.1.2 Hintergründe und Studienpopulation

Die Handröntgenbilder stammen von gesunden Kindern der Harpenden Growth Study, einer longitudinalen Studie, in deren Rahmen zwölf Jahre lang alle sechs Monate ein Röntgen der Probandinnen/Probanden angefertigt wurde (18). Anhand dieser Bilder wurde der Ablauf der Entwicklung in Stadien eingeteilt, die bei allen Kindern zu finden sind. Die Stadien sind unabhängig von Geschlecht und ethnischer Herkunft und wurden sowohl als Illustrationen abgebildet als auch verbal beschrieben, um sie möglichst genau zu definieren. Um das SA zu bestimmen, wird einzelnen Knochen am Handröntgen eines der acht oder neun Stadien zugewiesen. Das Stadium ergibt einen Score, der bei Mädchen und Knaben unterschiedlich sein kann, obwohl sich der Knochen im selben Stadium befindet. Die Scores der einzelnen Knochen werden addiert und ergeben einen Wert, der die Skelettreife quantitativ beschreibt, genannt Skeletal Maturity Score (SMS). Den SMS eines Kindes kann man ähnlich wie Größe und Gewicht mittels Perzentilenkurve mit dem anderer Kinder vergleichen und so Rückschlüsse ziehen, wie weit das Kind entwickelt ist (18). Die Referenzwerte für diese Kurven stammen in TW3 von drei verschiedenen Studien (18, 36, 37). Diese wurden im Zeitraum von 1969 bis 1995 durchgeführt und umfassten eine große belgische, eine spanische und eine amerikanische Population aus Texas. Röntgenbilder der ersten und dritten Gruppe wurden von nur je einer Untersucherin/einem Untersucher ausgewertet (18, 36). Die Referenzwerte von TW2 und TW3 unterscheiden sich, sodass in manchen Altersgruppen ca. ein Jahr Differenz vorliegt. Die Werte aus TW3 ähneln jenen anderer Populationen aus derselben Zeit mit vergleichbarem sozioökonomischem Hintergrund, beispielsweise Japan oder Nordamerika. Auch die Daten von Greulich und Pyle unterscheiden sich von TW3 nur wenig. Der Unterschied im SA ist mit 0,1-0,2 Jahren gering, obwohl die beiden Studien nicht zur gleichen Zeit durchgeführt wurden (2, 18). Der SMS ist zwischen den Populationen unterschiedlich, jedoch sind Vergleiche damit einfacher als mit dem SA.

Möchte man von dem SMS auf das SA schließen, muss man lediglich auf der Perzentilenkurve jene Stelle an der 50. Perzentile finden, an der der SMS mit dem der Probandin/des Probanden übereinstimmt und dann das CA ablesen, da dieses im Normalfall dem SA entspricht (18).

Es hat sich gezeigt, dass es günstiger ist, zwei getrennte Scores einzuführen: Einen für Handwurzelknochen und einen für Radius, Ulna, Mittelhand- und Fingerknochen (18). Ersterer ist der Karpalscore, letzterer wird als RUS-Score (Radius, Ulna and Short bone score) bezeichnet. Die beiden Scores variieren je nach Alter an Aussagekraft. Die Karpalknochen beispielsweise liefern ab einem gewissen Alter nur mehr ungenaue Information zur Entwicklungsgeschwindigkeit. Zusätzlich dazu wurde jedem einzelnen Knochen je nach Aussagekraft eine unterschiedliche Gewichtung zugeteilt, die den Score beeinflusst. Da zum Beispiel die Fingerknochen aufgrund ihrer Anzahl zusammen das Ergebnis verfälschen würden, gab man ihnen eine geringere Gewichtung (18). Werden sowohl die Knochen des RUS-Score als auch des Karpalscore beurteilt, spricht man von der 20-Knochen-Methode („20-bone method“)(38).

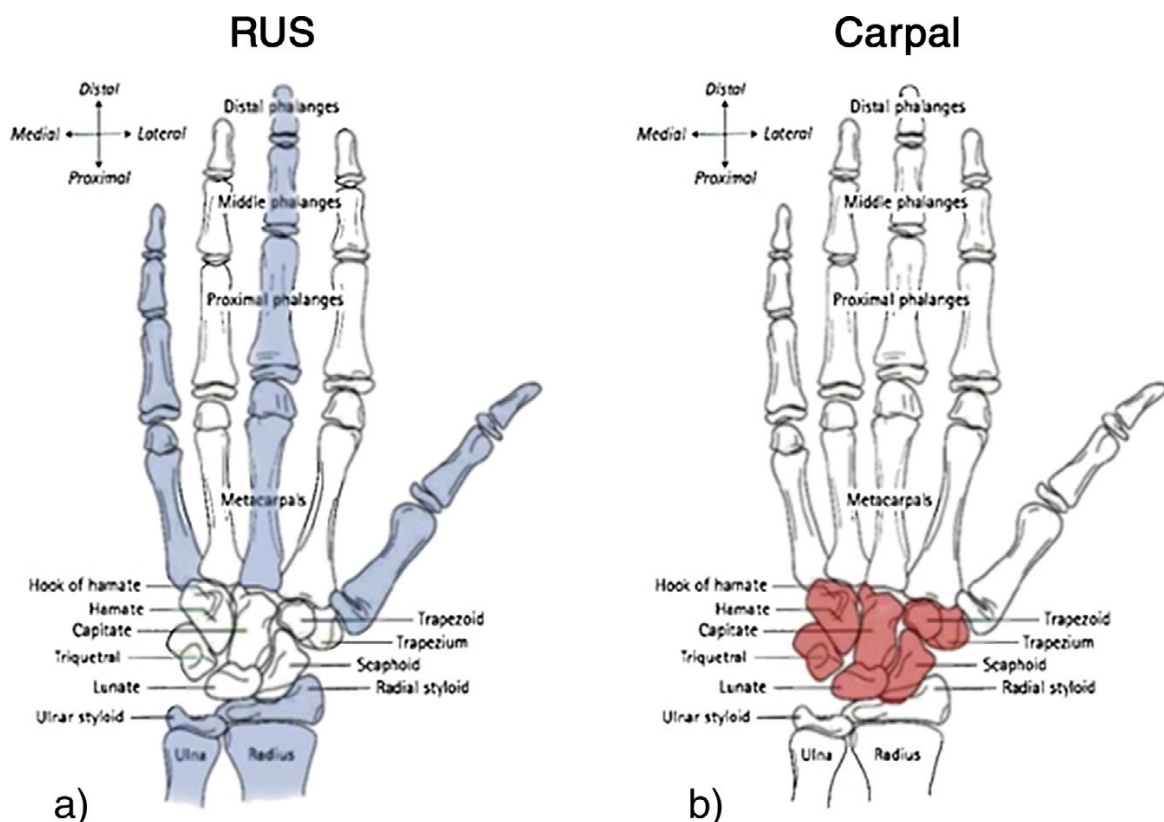


Abbildung 1: Knochen der beiden Scores: Für den RUS-Score werden Radius, Ulna sowie Metakarpalknochen und Phalangen des ersten, dritten und fünften Fingers beurteilt. Der Karpalscore setzt sich aus den Scores der Handwurzelknochen (exklusive Os pisiforme) zusammen (entnommen aus (38) mit Erlaubnis von Elsevier).

1.2.1.2.3 TW3 im Vergleich

Vorteil der TW3-Methode gegenüber anderen Verfahren ist, dass jeder relevante Knochen einzeln beurteilt wird. Das ist bei großen, meist durch pathologische Vorgänge verursachten Reifeunterschieden der Knochen von Bedeutung (39). Atlasvergleichsmethoden berücksichtigen dies weniger. Weiters ist es möglich, die Methode als Computerprogramm zu verwenden (18). Allerdings ist die TW3-Methode sehr zeitaufwendig: Sogar erfahrene Untersucherinnen/Untersucher benötigen für einen Befund vier- bis fünfmal so viel Zeit wie mit der Methode nach Greulich und Pyle (39). Auf eine korrekte Lagerung der Hand während der Aufnahme des Röntgenbildes ist zu achten, da es sonst zu fehlerhaften Befunden kommen kann (40). Werden die Epiphysenfugen von Radius und Ulna falsch beurteilt, kann dies besonders gravierende Auswirkungen auf den Befund der gesamten Hand haben, da diese Stellen im Score ein höheres Gewicht haben. Der Unterschied zwischen zwei Stadien ist besonders gegen Ende des Wachstums groß, weshalb für forensische Indikationen feinere Abstufungen wünschenswert sind (39). Für klinische Indikationen sind die Limitationen der Methode zu beachten (40).

In einer Studie von Schmidt et al. werden die Methoden TW2 und TW3 miteinander verglichen (39). Dabei wurden zwischen 1986 und 2002 aufgenommene Röntgenbilder von 44 Mädchen und 48 Knaben aus Deutschland im Alter von 12-16 Jahren verwendet. Beide Methoden zeigten zwar eine starke Korrelation zwischen SA und CA, TW2 überschätzte das Alter jedoch teilweise stark. Besonders bei forensischen Fragestellungen ist eine solche Überschätzung nicht zulässig. Das mittels TW3 erhobene SA in der Altersgruppe von 14 bis 16 Jahren unterschied sich vom Mittelwert des CA nur gering, die Abweichungen betragen -0,4 bis +0,2 Jahre. Die Autoren schlussfolgerten, dass TW3 daher für forensische Indikationen empfehlenswert ist (39).

Eine weitere Studie verglich TW2, TW3 und die Handskeletalterbestimmung nach Greulich und Pyle in einer Population bestehend aus 46 adoptierten Mädchen, die frühzeitig die Pubertät erreichten (41). Verglichen wurden die nach diesen Methoden vorausgesagten Erwachsenengrößen mit den tatsächlichen endgültigen Größen, um herauszufinden, welche Methode für die Selektion der Patientinnen für eine kombinierte Hormontherapie am geeignetsten ist. Es zeigte sich, dass die Methode nach Greulich und Pyle dazu besser geeignet war als TW3 und TW2 (41).

In einer australischen Studie wurden Röntgenbilder von 360 Probandinnen/Probanden im Alter von 0-24,9 Jahren mittels TW2- und TW3-Methode befundet (38). Die genauesten SA-Schätzungen waren bei Mädchen der RUS-Score der TW3-Methode und bei Knaben der Karpalscore von TW2. Die Studie ergab, dass sich die Methoden für forensische Fragestellungen bei jüngeren australischen Kindern eignen. Da aber das letzte Stadium der Skelettreife meist schon vor Vollendung des 15. Lebensjahres erreicht wurde, sind die Methoden nicht geeignet, um Aussagen über Erreichen der Volljährigkeit zu treffen (38).

1.2.1.3 Die FELS-Methode

1988 stellten Chumlea, Roche und Thissen die FELS-Methode vor (19). Zwischen 1932 und 1972 wurden im Rahmen der Fels Longitudinal Study 13823 Röntgenbilder der linken Hand von Mädchen und Knaben gesammelt. Die Probandinnen/Probanden wurden bis über das 18. Lebensjahr hinaus begleitet, wobei im ersten Lebensjahr fünf Röntgenbilder aufgenommen wurden. Bis zum 18. Geburtstag wurden die Untersuchungen in Intervallen von sechs Monaten durchgeführt, danach alle zwei Jahre. Ähnlich wie bei der Methode nach Greulich und Pyle wurde auch hier nach „maturity indicators“ gesucht (2, 19). Es wurde allerdings versucht, diese genau festzulegen, einzuteilen und zu objektivieren. Von 130 radiologischen Merkmalen, die auf ihre Eignung als „maturity indicator“ überprüft wurden, wurden 98 ausgewählt. Davon wurden 85 in mehrere Stadien unterteilt. Die restlichen 13 Merkmale sind metrische Indikatoren, die am Röntgenbild ausgemessen werden. Anhand der vorliegenden Daten aus der Fels Longitudinal Study wurde errechnet, in welchem CA ein Indikator mit einer Wahrscheinlichkeit von 50% vorhanden bzw. in einem bestimmten Stadium ist (19). Soll nun das SA bestimmt werden, überprüft man das Vorhandensein der Indikatoren am Röntgenbild der Patientin/des Patienten. Welche und wie viele Indikatoren untersucht werden, ist von CA und Geschlecht abhängig (19). Die Kombination der vorhandenen Indikatoren wird mit der Studienpopulation verglichen, um das CA jener Probanden, die dieselben Indikatoren und somit eine ähnliche Skelettreife aufweisen, zu ermitteln. Dieses CA ist das SA der untersuchten Patientin/des untersuchten Patienten. 98 Merkmale, von denen 85 in verschiedenen Stadien vorliegen können, ermöglichen eine hohe Anzahl an Kombinationen. Die Studienpopulation hätte noch wesentlich größer sein müssen, um alle möglichen Kombinationen abzudecken (42). Daher wird ein statistisches Verfahren verwendet, das die Schätzung des SA mit der untersuchten Studienpopulation erlaubt. Das Vorgehen wird

durch ein Computerprogramm erleichtert, das nicht nur das geschätzte SA liefert, sondern auch den mittleren Fehler, was eine Besonderheit dieser Methode ist (42).

Die FELS-Methode wurde mit der Handskelettalterbestimmung nach Greulich und Pyle und der TW3-Methode verglichen (19). Dazu wurden die Röntgenbilder aus der Fels Longitudinal Study verwendet. Größere Unterschiede zeigten sich zwischen FELS-Methode und TW3-Methode (19). Auch an einer italienischen Population wurden dieselben Methoden getestet (43). Die verwendeten Röntgenbilder entstanden von 1985 bis 1990. Die Methode nach Greulich und Pyle schätzte das SA sehr nah am CA, während die FELS-Methode zu hohe Schätzungen lieferte. Die Abweichungen wurden mit ansteigendem Alter der Probandinnen/Probanden größer, die Ergebnisse lagen jedoch laut Vignolo et al. noch in einem akzeptablen Bereich und erwiesen sich als besser reproduzierbar als mittels anderer Methoden (43).

1.2.2 Andere Röntgen-basierte Methoden der Altersbestimmung

1.2.2.1 Dental Age Assessment

Auch die Entwicklung der Zähne kann zur Altersbestimmung herangezogen werden. Besonders in Forensik, Zahnheilkunde und Kieferorthopädie sind solche Methoden weit verbreitet (44). Vorteilhaft dabei ist, dass die Entwicklung der Zähne nur selten von Ernährung und nur wenig von endokrinen Faktoren beeinflusst wird und erst spät abgeschlossen ist. Allerdings gibt es viele andere Faktoren, die sich auf die Zahnentwicklung auswirken können (44). Eine häufig eingesetzte Methode ist die Zahnaltersbestimmung nach Demirjian, für die ein Panoramaröntgen notwendig ist (45). Dabei werden sieben Zähnen im linken Unterkiefer Stadien zugewiesen, die einen Score ergeben. Dieses Vorgehen ähnelt der Methode nach Tanner und Whitehouse (18, 45).

1.2.2.2 Risser-Zeichen

Auch der Beckenkamm durchläuft in der Entwicklung des menschlichen Körpers einige Veränderungen und kann daher als Indikator für die Skelettreife dienen (46). Die Ossifikation der Beckenkammepophysen wurde von Risser in sechs Stadien eingeteilt (46, 47). Seit 1958 wird das Risser-Zeichen zur Abschätzung des SA und damit auch des Wachstumspotenzials der Wirbelsäule verwendet (46-48). Da eine Korrelation zwischen Risser-Zeichen und Wirbelsäulenwachstum vorhanden ist, spielt es noch heute in der

Behandlung der Skoliose im Kindes- und Jugendalter eine wichtige Rolle (46-49). Das Risser-Zeichen wird anhand einer Röntgenaufnahme des Beckenkammes beurteilt. Aufgrund der anatomischen Nähe zur Wirbelsäule ist der Beckenkamm an einem Röntgenbild mit anterior-posteriorem Strahlengang der gesamten Wirbelsäule ebenso ersichtlich und das Stadium kann aus demselben Bild bestimmt werden (1). Das ist besonders vorteilhaft, da bei Patientinnen/Patienten mit Skoliose ohnehin ein solches Röntgenbild aufgenommen wird (47). Die Strahlenbelastung und der Aufwand eines zusätzlichen Handröntgens werden so erspart. In den vielen anderen Fällen, in denen noch kein Röntgen vorliegt und eine genaue SA-Bestimmung notwendig ist, ist dem Handröntgen der Vorzug zu geben (1). Grund dafür ist, dass Methoden zur SA-Bestimmung mittels Handröntgen in der Regel genauer sind als das Risser-Zeichen (47, 49). Das Risser-Zeichen korreliert stärker mit dem SA als mit dem CA (1). Es zeigte sich, dass die Fusion der Beckenkammepiphyse ein guter Indikator für das Ende des Wirbelsäulenwachstums ist (46, 49). Das bedeutet auch, dass ab diesem Zeitpunkt das Risiko einer Progression einer adoleszenten Skoliose kaum mehr gegeben ist (47, 49). Die Ossifikation der Beckenkammepiphyse ist bei Mädchen im Alter von 14 Jahren, bei Knaben im Alter von 16 Jahren abgeschlossen. Akzeleriert kann sie beispielsweise auf der höheren Seite des Beckens bei einem Beinlängenunterschied sein (46). Alles in allem ist das Risser-Zeichen ein schnell verfügbarer Indikator, der auch heute noch in der Orthopädie regelmäßig Anwendung findet.

1.2.2.3 Skeletalterbestimmung am Knie

Auch Röntgenbilder des Kniegelenks können zur Altersbestimmung herangezogen werden. 1969 veröffentlichten Pyle und Hoerr zu diesem Zweck einen Atlas mit Röntgenbildern des Kniegelenks ähnlich dem Handröntgenatlas von Greulich und Pyle (2, 50). Er wird zwar heute noch verwendet, jedoch deutlich seltener als die Handskeletalterbestimmung nach Greulich und Pyle (2, 17). Es besteht Grund zur Annahme, dass es Unterschiede zwischen den mittels Knie- und Handröntgenbildern durchgeführten Altersschätzungen gibt (51, 52). Dies widerspricht der Aussage von Greulich und Pyle, das Skelett würde sich als Einheit überall gleichmäßig entwickeln (2). Für Fragestellungen, die das Wachstum der unteren Extremität betreffen, erscheint es daher sinnvoll, das SA an dieser Lokalisation zu bestimmen.

1.2.3 MRT-basierte Methoden

Im Sinne der Strahlenhygiene wurden im Laufe der Zeit Methoden zur Bestimmung des SA entwickelt, die ohne Röntgenaufnahmen auskommen (53-63). Urschler et al. zeigten, dass die bereits bestehenden Methoden nach Greulich und Pyle und TW2 auch an MRT-Aufnahmen anwendbar sein könnten (53). Terada et al. führten eine ähnliche Studie durch und gelangten zum selben Schluss (55).

Edmonds et al. befassten sich mit den morphologischen Veränderungen in MRT-Aufnahmen, die das Knie im Alter von 4 bis 18 Jahren durchläuft (54). Es wurden Längenmessungen knöcherner und knorpeliger Strukturen durchgeführt, um altersspezifische Normwerte zu ermitteln. Die Maße fast aller gemessenen Strukturen korrelierten mit dem CA (54). Pennock et al. erstellten 2017 einen Atlas mit MRT-Aufnahmen des Knies (60). Dazu wurden über 800 MRT-Datensätze von Patientinnen/Patienten im Alter von 2 bis 19 Jahren verwendet. Altersspezifische Merkmale und Veränderungen wurden dokumentiert. So zeigte sich, dass die Patella bei Kleinkindern die größte Rolle für die Altersschätzung spielt, während bestimmte Veränderungen der Tibia besonders im Alter von 6 bis 12 Jahren auftreten. Später gewinnen Femur und Fibula in der Altersschätzung an Bedeutung. Der Atlas wurde anhand weiterer 323 MRT-Aufnahmen validiert, was eine starke Korrelation zwischen dem geschätzten SA und dem CA ergab (60).

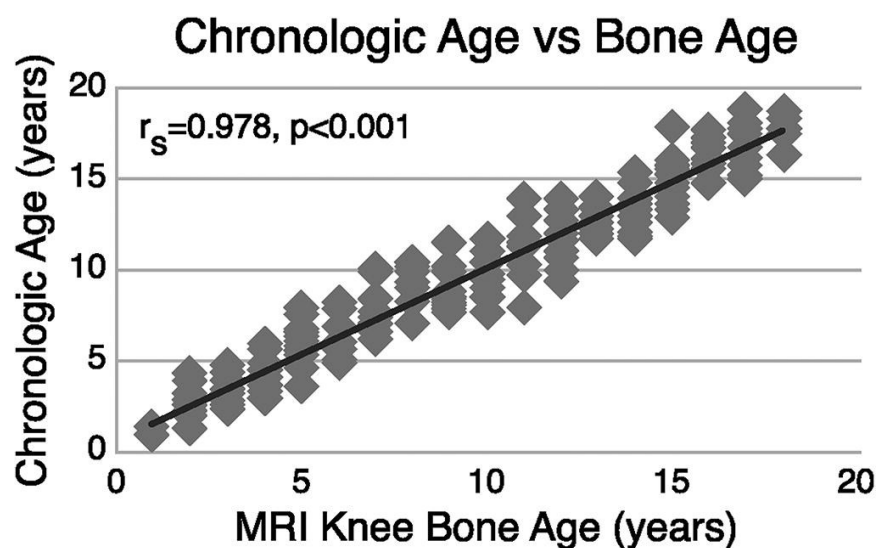


Abbildung 2: Korrelation zwischen CA und mittels MRT des Kniegelenks bestimmtem SA (entnommen aus (60) mit Erlaubnis von Wolters Kluwer Health, Inc.).

In einer retrospektiven Studie untersuchten Dedouit et al. 290 MRT-Aufnahmen des Knies von Patientinnen/Patienten im Alter von 10 bis 30 Jahren. Es wurde ein Stagingssystem bestehend aus fünf Stadien entwickelt, das die Reife der Wachstumsfugen des distalen Femur und der proximalen Tibia klassifiziert. In der Validierung des Stagingystems zeigten sich nur geringe Intra- und Interobserverabweichungen (56). Ekizoglu et al. prüften die Methode an einer türkischen Population und fanden eine signifikante Korrelation zwischen Ossifikationsstadium und Alter (64). Krämer et al. beurteilten die distale Femurepiphyse an MRT-Aufnahmen von ca. 300 Probandinnen/Probanden und teilten sie nach ihrem Entwicklungsstand in Ossifikationsstadien ein. Dazu wurde eine Kombination bereits bestehender Klassifikationen verwendet (58). Es konnte bestätigt werden, dass die Stadien auch in einer anderen Population, die nur aus männlichen Probanden bestand, mit dem CA korrelieren (65). In einer weiteren Studie wurde auch die proximale Tibiaepiphyse untersucht. Wiederum wurde eine modifizierte Klassifikation vorgenommen (57). Fan et al. untersuchten retrospektiv Knie-MRT-Aufnahmen einer chinesischen Population und teilten die Befunde in modifizierte Stadien nach Krämer et al. ein (57-59). Die MRT-Aufnahmen wurden mit Röntgenaufnahmen des Knies verglichen, die am selben Tag aufgenommen worden waren. Sowohl das aus den Röntgenaufnahmen erhobene SA als auch das MRT-SA zeigten eine starke Korrelation zum CA, wobei das MRT-SA weniger Abweichungen aufwies (59).



Abbildung 3: Vergleich der Bildmodalitäten: Verschiedene Stadien der Entwicklung des Kniegelenks in MRT-Aufnahmen (oben) und auf Röntgenbildern (unten)(entnommen aus (59) mit Erlaubnis von Elsevier).

In einer rezenten, prospektiven Studie von Auf der Mauer et al. wurden über den Zeitraum von zwei Jahren drei MRT-Untersuchungen des Kniegelenks von 40 männlichen Probanden aus Deutschland durchgeführt (61). Das Alter der Probanden lag zwischen 14 und 21 Jahren. Zusätzlich zu den MRT-Untersuchungen wurden Körpergröße und -gewicht sowie die Beinlänge erhoben. Die MRT-Befunde wurden in drei Ossifikationsstadien eingeteilt. Auch diese Studie ergab eine Korrelation zwischen CA und Stadium der Ossifikation der Knochen des Kniegelenks sowie nur geringe Interobserverabweichungen (61).

Methoden zur SA-Bestimmung mittels MRT wurden in den letzten Jahren verstärkt erforscht und zeigen vielversprechende Resultate. Es existieren auch Methoden zur Altersschätzung mittels Ultraschall (66-68).

1.2.4 Automatisierte Methoden

Um die subjektive Komponente einer SA-Bestimmung zu eliminieren und den Zeitaufwand der Beurteilung zu minimieren, wurden Computerprogramme entwickelt, die das SA automatisch berechnen. Durch den technischen Fortschritt ergeben sich immer mehr Möglichkeiten in diesem Bereich. Schon vor der TW3-Methode wurde das System CASAS (Computer-Assisted Skeletal Age Scoring System) vorgestellt (18, 69). Dieses war jedoch sehr zeitaufwendig und benötigte manuelle Korrekturen der Ergebnisse (70). Die menschliche Erfahrung und Intuition durch ein komplexes System aus Regeln zu ersetzen stellte sich als weniger effizient heraus, als man zu Beginn angenommen hatte. So wichen die anfänglichen regelbasierten Ansätze sogenannten Neural Networks und Methoden des Deep Learning, die im folgenden Kapitel näher beschrieben werden (70, 71). 1995 wurde bereits ein einfaches Neural Network entwickelt, das in 47% der getesteten Fälle ein genaueres SA berechnete als die Radiologin/der Radiologe (72).

2008 wurde die Software BoneXpert veröffentlicht (70). BoneXpert erkennt die Umrisse von 15 Knochen der Hand selbständig. Verschiedene Werte der Knochendichte, -form und -textur werden erhoben. Anhand dieser Eigenschaften wird ein intrinsisches SA errechnet, das in einem letzten Schritt in ein SA nach Tanner und Whitehouse oder Greulich und Pyle umgewandelt wird. Die Daten für die Entwicklung der Software stammen von belgischen und dänischen Studienpopulationen. Das durch BoneXpert errechnete SA nach Greulich

und Pyle entspricht größtenteils dem manuell erhobenen SA, allerdings lehnt das System Bilder mit niedriger Qualität ab. BoneXpert wurde an zahlreichen Studienpopulationen validiert. Die Genauigkeit wurde als „root mean square deviation“ (RMSD, Wurzel der mittleren quadratischen Abweichung) von der manuellen Handskelettalterbestimmung nach Greulich und Pyle definiert. Die erste Validationsstudie ergab eine RMSD von 0,71 Jahren (8,52 Monate). In der Tübingen-Studie lag die RMSD bei 0,72 Jahren (8,64 Monate). Eine weitere Studienpopulation bestand aus 1103 Kindern unterschiedlicher ethnischer Herkunft. Die RMSD betrug 0,59 Jahre (7,08 Monate). Es konnten keine signifikanten Unterschiede zwischen den Ergebnissen von Kindern unterschiedlicher ethnischer Herkunft festgestellt werden, weshalb die Autoren schlussfolgerten, dass die Methode von der Ethnie unabhängig ist. In allen genannten Studien wurden die SA-Schätzungen von BoneXpert an dem SA nach Greulich und Pyle gemessen (70).

Ein Deep Learning-System von Lee et al. ermöglicht durch den Einsatz eines Convolutional Neural Networks (CNN) die vollautomatische SA-Bestimmung (73).

Die meisten bisher entwickelten automatisierten Methoden verwenden Röntgenbilder der Hand und/oder orientieren sich an bereits bestehenden manuellen Methoden (70, 71, 73). Stern et al. stellten jedoch 2014 eine automatisierte Methode vor, die MRT-Aufnahmen der linken Hand zur SA-Bestimmung verwendet. Es konnten Ergebnisse erzielt werden, die denen des Radiologen sehr ähnlich waren (74).

1.3 Voraussage der Erwachsenengröße

Die Voraussage der Körpergröße zu Wachstumsabschluss ist von großem Interesse bei der Planung korrigierender Eingriffe (1). Oft ist es notwendig, Eltern und Kind in Bezug auf die Körpergröße des Kindes zu beraten. Eine verlässliche Vorhersage der Erwachsenengröße kann die Notwendigkeit gewisser diagnostischer Schritte oder einer Therapie aufzeigen (75). Aber auch zur Überwachung des Therapieerfolgs werden Methoden zur Voraussage der Erwachsenengröße eingesetzt (11). Häufig wird das SA als Basis für die Voraussage der verbleibenden Wachstumslänge herangezogen (12, 18, 76, 77). Ungenauigkeiten und Fehler in der Schätzung des SA pflanzen sich daher in den zweiten Schritt, die Berechnung der Erwachsenengröße, fort. Ebenso ist zu berücksichtigen, dass durch die Wahl einer für die untersuchte Population nicht gültigen Formel zur Berechnung der endgültigen Größe Fehler entstehen können (78). Je weiter ein Kind vom Wachstumsabschluss entfernt ist, desto schwieriger ist die genaue Voraussage der endgültigen Körpergröße (1). Das Längenwachstum eines Kindes wird vor allem durch genetische Gegebenheiten bestimmt. Aber auch Stoffwechselerkrankungen, neuromuskuläre Erkrankungen und Umweltfaktoren wie beispielsweise die Ernährung haben einen Einfluss auf die Erwachsenengröße (1). Des Weiteren wirken Zug und Druck mechanisch stimulierend bzw. hemmend auf das Knochenwachstum.

1.3.1 Anwendungsbereiche

In der Orthopädie hat die Voraussage der Erwachsenengröße zahlreiche Anwendungsbereiche. Zur zeitlichen Planung korrigierender orthopädisch-chirurgischer Eingriffe am wachsenden Skelett ist sie unverzichtbar (1).

In der Behandlung von Beinlängendifferenzen und Achsfehlstellungen im Wachstumsalter gilt es, den optimalen Zeitpunkt für eine chirurgische Intervention zu finden (79). Es muss abgeschätzt werden, wie viel Längenwachstum noch zu erwarten ist. Darüber hinaus ist auch die zeitliche Vorhersage des Wachstumsabschlusses des Individuums von Interesse. Mit dieser Information lassen sich auch Progression und die zu erwartende resultierende Beinlängendifferenz nach Abschluss des Wachstums abschätzen (12). Um Patientinnen/Patienten mit Achsfehlstellungen oder Beinlängenunterschieden die aufwendige Korrekturosteotomie oder Beinverlängerungsverfahren zu ersparen, kann im

Wachstumsalter eine weniger invasive Epiphyseodese oder Hemiepiphyseodese im Sinne einer Wachstumslenkung oder Wachstumsbremsung erfolgen (14, 79, 80). Auf die Gesundheit der pädiatrischen Patientinnen/Patienten wirkt sich dies außerordentlich positiv aus, da durch solche Eingriffe präarthrotische Deformitäten, die frühzeitig zum schmerzhaften Gelenksverschleiß führen, bereits im Wachstumsalter vermieden werden können (79). Wachstumslenkende Eingriffe sind jedoch nur möglich, solange das Wachstum nicht abgeschlossen ist, daher darf zur Vermeidung einer Unterkorrektur der geeignete Zeitpunkt nicht verpasst werden. Aber auch von der verfrühten Durchführung eines solchen chirurgischen Eingriffes ist abzuraten, da in diesem Fall eine Überkorrektur droht (79).

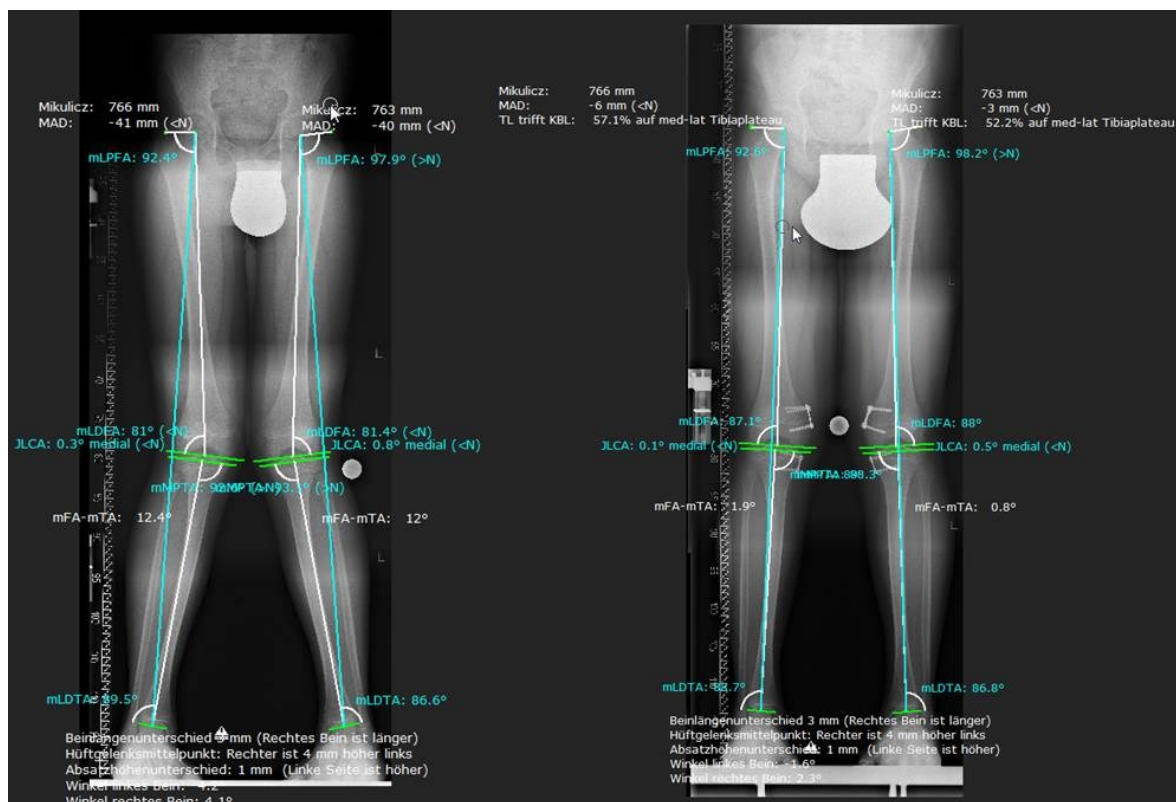


Abbildung 4: Orthopädische Korrekturplanung: Zur Planung der wachstumslenkenden Achskorrektur bei Genua valga mittels temporärer Hemiepiphyseodese (Eight-Plates) ist die Kenntnis des verbleibenden Wachstumspotentials unerlässlich (von der Abteilung für Kinderorthopädie der Universitätsklinik für Kinder- und Jugendchirurgie zur Verfügung gestellt).

Korrekte zeitliche Planung ermöglicht die Erreichung des Zieles, dass bei Wachstumsabschluss der Beinlängenunterschied oder die Achsfehlstellung korrigiert ist (14). Ist ein Extremitäten-verlängerndes Verfahren mittels eines Fixateur Externe oder eines intramedullären Systems notwendig, muss die erforderliche Verlängerungsstrecke

berechnet werden (79). Bestimmung des SA, Voraussage der zu erwartenden Erwachsenengröße oder Extremitäten- bzw. Knochenlänge sowie die verbleibende Zeit bzw. das verbleibende Wachstumspotential bis zum Wachstumsende sind für eine solche orthopädische Korrekturplanung essentiell.

Ein weiterer Bereich, in der SA und Voraussage der Erwachsenengröße zum Einsatz kommen, ist die Therapie der Skoliose im Kindes- und Jugendalter. Dabei entscheidet die Bestimmung des SA über den erfolgreichen Einsatz eines Korsetts (Stützmieters) zur Korrektur der Wirbelsäulenverkrümmung (48). Nicht nur die konservative Korsetttherapie ist auf das SA angewiesen, auch vor korrigierenden chirurgischen Eingriffen muss das SA bestimmt werden (81). Prinzipiell ist der Zeitpunkt derartiger Korrekturen so spät wie möglich zu wählen. Ziel der Behandlung ist die Auskorrektur zu Wachstumsende, Über- oder Unterkorrektur sollen vermieden werden. Wird der Zeitpunkt der Operation zu früh gewählt, riskiert man durch das weitere Knochenwachstum ein Wiederauftreten der Verkrümmung der Wirbelsäule im Sinne eines Crankshaft-Phänomens (81). Die Wachstumsprognose ist auch wichtig, um die Wahrscheinlichkeit einer Progredienz der Krümmung im verbleibenden Wachstum einschätzen. Die Therapie wird dementsprechend angepasst (48). Das höchste Risiko einer Progredienz besteht während des präadoleszenten Wachstumsschubes (82).

Auch in der Endokrinologie muss die voraussichtliche Erwachsenenkörperhöhe oft erhoben werden. Groß- oder Kleinwuchs, adrenogenitales Syndrom, Hypo- oder Hyperthyreose und Hormonersatztherapie bei Somatotropinmangel sind nur einige Beispiele, bei denen für Diagnose und/oder Überwachung des Therapieerfolges eine Voraussage der Erwachsenengröße erforderlich ist. Wie bereits beschrieben wird dafür meist das SA als Basis verwendet (78).

Eine ungewöhnlichere Möglichkeit der Anwendung von Größenvorhersagen besteht in der Sportwissenschaft. Für junge Athletinnen/Athleten ist ihre voraussichtliche Erwachsenengröße von Interesse, wenn in ihrer Sportart eine bestimmte Größe notwendig oder von Vorteil ist. Ihre sportliche Karriere kann von den Ergebnissen der Vorhersage beeinflusst werden (75).

1.3.2 Methoden zur Voraussage der Erwachsenenengröße

Einige Methoden zur Voraussage der Erwachsenenengröße entstanden auf Grundlage der dazugehörigen Methode der SA-Bestimmung (18, 76, 77). So liefert die TW3-Methode auch eine Formel zur Voraussage der Erwachsenenengröße und die dazu notwendigen Daten. Die Vorhersage verwendet nur den RUS-Score anstelle des SMS und die Konstanten a und b, die altersabhängig sind.

$$\text{Erwachsenengröße} = \text{aktuelle Größe} + a \times \text{RUS-Score} + b$$

Die Autoren schlagen vor, die erwartete Größe z.B. als „wahrscheinlichste Größe 160 cm und wahrscheinliche Grenzen 157 cm bis 163 cm“ anzugeben (18). Die Handskelettalterbestimmung nach Greulich und Pyle dient als Basis für die Voraussage der Erwachsenenstehhöhe nach Bayley und Pinneau (2, 76). Allerdings wurden auch andere Methoden entwickelt, von denen nicht alle die Kenntnis des SA als Voraussetzung haben (13).

1.3.2.1 Voraussage der Erwachsenenstehhöhe nach Bayley und Pinneau

Die Voraussage der Erwachsenenstehhöhe nach Bayley und Pinneau wurde um 1950 entwickelt und 1959 überarbeitet (76). Dabei handelt es sich um eine einfach durchzuführende, weit verbreitete Methode. Sie basiert auf der Überlegung, dass ein enger Zusammenhang zwischen dem SA nach Greulich und Pyle und dem prozentuellen Anteil der Erwachsenenengröße, den das Kind zu dem Zeitpunkt der Untersuchung erreicht hat, besteht. Die Daten der Methode stammen von einer Population aus 192 gesunden Kindern aus der Berkeley Growth Study, die im Alter von 8 bis 18 Jahren alle sechs Monate untersucht wurden. Als die Größe bei Abschluss des Wachstums bekannt war, wurde diese als 100% angenommen und für alle vorherigen Messungen der sogenannte „per cent of mature height“-Wert (PMH) berechnet. Dieser wurde auf mehreren Tabellen für Mädchen und Knaben dargestellt. Die Erwachsenenengröße kann durch Division der Größe des Kindes durch den PMH für das SA errechnet werden (76). Ein eindeutiger Nachteil dieser Methode ist das Fehlen akkurater Daten für Kinder unter acht Jahren (83).

1.3.2.2 Die Roche-Wainer-Thissen-Methode

Die Roche-Wainer-Thissen-Methode (RWT-Methode) wurde von den gleichnamigen Autoren 1975 veröffentlicht (83). Mittels der RWT-Methode kann die Körperlänge, die das Kind im Alter von 18 Jahren erreichen wird, vorausgesagt werden. Dies ist mit nur einem Messzeitpunkt, zu dem vier Parameter erhoben werden, möglich und daher schnell verfügbar (83).

Die für die Voraussage der Erwachsenengröße benötigten Unterlagen sind zwei Tabellen, eine für Mädchen und eine für Knaben, sowie Angaben zu den Fehlergrenzen (83). Die Tabellen zeigen die unterschiedliche Gewichtung der vier zu erhebenden Parameter für die einzelnen Altersgruppen. Das Alter ist in Abstände von je drei Monaten unterteilt und reicht bei Mädchen bis 14 Jahre, bei Knaben bis 16 Jahre. Die Werte der unterschiedlichen Gewichtungen ergaben sich aus den Daten der Studienpopulation. Dabei handelte es sich um gesunde, kaukasische Probandinnen/Probanden aus Ohio, die am Fels Research Institute alle sechs Monate untersucht wurden. Die Körpergröße im Liegen („recumbent length“), das Gewicht, die Größe der Eltern und das SA wurden erhoben. Das SA wurde mittels Handskelettalterbestimmung nach Greulich und Pyle, Fußskelettatlas von Hoerr et al. und Kniegelenksatlas von Pyle und Hoerr bestimmt. Pro Geschlecht und Altersgruppe lagen ca. 100 Datensätze vor. Es herrschte Unsicherheit, welche Methode zur Bestimmung des SA für die Voraussage der Erwachsenengröße am besten geeignet wäre. Daher wurden statistische Analysen durchgeführt, um die Korrelationen zwischen den SA-Werten der einzelnen Knochen zu ermitteln. Von den 78 Parametern, die überprüft wurden, wurden die 18 am besten geeigneten ausgewählt. Im nächsten Schritt wurden diese in verschiedenen Kombinationen getestet und auf vier reduziert. Diese vier Parameter sind die Körpergröße im Liegen, das Körpergewicht, die Körpergröße der Eltern und das SA nach Greulich und Pyle. Sie erwiesen sich unabhängig von Alter und Geschlecht als für die Voraussage der Erwachsenengröße geeignet. Wie bereits beschrieben genügt ein Arztbesuch, um diese Werte zu erheben, was in der Anwendung von Vorteil ist. Man entschied sich für die Körpergröße in liegender Position anstelle der Stehgröße, da sie auch bei Säuglingen und Kleinkindern messbar ist. Die RWT-Methode ist nur anwendbar, solange nicht mehr als die Hälfte der Handwurzelknochen vollständig ausgereift sind (83).

Die vier erhobenen Werte werden mit den zugehörigen tabellarischen Multiplikatoren, die die Aussagekraft widerspiegeln, multipliziert (83). Die Ergebnisse werden addiert und stellen schon den ersten Richtwert dar, der in cm angegeben wird (z.B. 170 cm). Anschließend werden die Fehlergrenzen für das Alter des Kindes abgelesen und angegeben, z.B. ± 5 cm. Die RWT-Methode liefert somit den oberen und unteren Grenzwert der Größe, die 90% der Kinder mit denselben Ausgangswerten haben werden (83). Die Aussage könnte beispielsweise lauten: Die Patientin/der Patient wird mit einer Wahrscheinlichkeit von 90% mit 18 Jahren eine Größe von 165 bis 175 cm haben.

Die RWT-Methode wurde an den Daten dreier großer longitudinaler Wachstumsstudien getestet und mit der Voraussage der Erwachsenenstehhöhe nach Bayley und Pinneau verglichen (83). Bei den Daten der Kinder, die am Fels Research Institute untersucht wurden, schnitt die RWT-Methode besser ab als die Methode nach Bayley und Pinneau. Da die RWT-Methode mit denselben Daten entwickelt wurde, war dies nicht weiter verwunderlich. Allerdings wies sie auch bei den weiblichen Probanden der Studie des Child Research Council in Denver und bei den männlichen Probanden der Studie der Harvard School of Public Health weniger Fehler auf (83).

1.3.2.3 Verbleibendes Wachstum der unteren Extremitäten nach Anderson

Viele Methoden für die Voraussage der Erwachsenengröße basieren auf Forschungsergebnissen von Anderson et al. (12, 13). Um die Progression von Beinlängenunterschieden abzuschätzen wurden Messungen der Länge von Femur und Tibia durchgeführt (12, 13).

Die Studienpopulation bestand aus 100 Kindern, davon 50 Mädchen und 50 Knaben (12). 51 der Kinder waren gesund, bei 49 lag ein Zustand nach Poliomyelitis vor. Bei der letzteren Gruppe wurden die Messungen am von der Paralyse nicht betroffenen Bein durchgeführt. Die Kinder wurden ab dem Alter von acht Jahren bis zum Abschluss des Wachstums der unteren Extremitäten einmal jährlich gemessen. Die Länge von Femur und Tibia wurde mittels Röntgenbild bestimmt. Auch wurden Körperlänge, Länge von Stamm, Extremitäten und Füßen und das SA nach Greulich und Pyle erhoben. Anderson et al. empfahlen, die Werte für das noch zu erwartende Wachstum immer gemeinsam mit dem SA und nicht mit dem CA anzugeben. Daher wurde das Längenwachstum der

Knochen immer in gleichmäßigen zeitlichen Abständen des SA gemessen, was Voraussagen präziser macht. Anhand weiterer Daten wurde auch errechnet, wie hoch der Anteil der einzelnen Epiphysen am Längenwachstum ist. Untersucht wurden dabei der distale Femur und die proximale Tibia. Mit den Tabellen und Darstellungen von Anderson et al. lässt sich das verbleibende Wachstum eines Beines bzw. von Femur und Tibia in einem bestimmten skelettalen Alter ablesen. In der Planung wachstumslenkender Eingriffe sind diese Informationen wertvoll (12).

1.3.2.4 Moseley Straight-Line Graph

Der Moseley Straight-Line Graph (M-SLG) wurde um 1977 für die Wachstumsvoraussage bei Beinlängenunterschieden entwickelt (84). Dazu wurden Daten von Anderson et al., die um 1964 publiziert wurden und Messwerte von Kindern ab einem Alter von einem Jahr enthielten, verwendet und grafisch dargestellt. Moseley modifizierte die Zeitachse, sodass sich die Werte, die normalerweise keine lineare Funktion ergeben, auf einer geraden Linie anordnen (13). Die Methode berücksichtigt die Wachstumsperzentile des Kindes und hilft, Veränderungen der Wachstumsgeschwindigkeit zu erkennen (84). Allerdings werden zwei oder mehr Messungen benötigt, die im Abstand von mindestens einem Jahr durchgeführt werden sollten, um genaue Ergebnisse zu erzielen (13). Der M-SLG wird bis heute für die entsprechenden Fragestellungen verwendet.

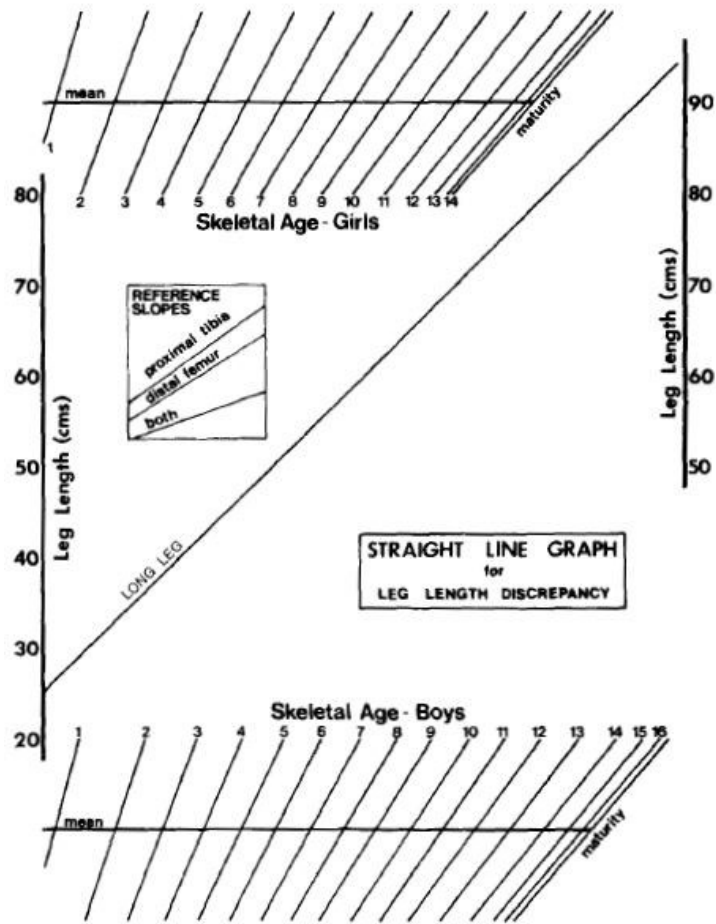


Abbildung 5: Moseley Straight-Line Graph (entnommen aus (84) mit Erlaubnis von Wolters Kluwer Health, Inc.).

1.3.2.5 Multiplier Method

Die von Paley et al. entwickelte Multiplier Method basiert auf denselben Daten, die auch Moseley verwendete, ergänzt mit Daten von Kindern im ersten Lebensjahr (13). Die Länge von Femur und Tibia bei abgeschlossenem Wachstum bzw. skelettaler Reife wurde für Mädchen und Knaben jeder Perzentile als L_m angegeben. Teilt man L_m durch die Länge des Knochens in einem bestimmten Alter L , erhält man einen Multiplikator M (13).

$$M = L_m/L$$

Dies wurde für alle Altersgruppen von einem Jahr bis zur Skelettreife und für den jeweiligen Knochen berechnet. Umgekehrt ist es bei Kenntnis der derzeitigen Knochenlänge L möglich, die endgültige Länge des Knochens L_m zu bestimmen. Der dazu notwendige altersspezifische Multiplikator M kann in einer Tabelle abgelesen werden (13, 85).

$$L_m = L \times M$$

TABLE I
FEMORAL MULTIPLIERS FOR BOYS BY PERCENTILE GROUP

Age (yrs.)	Multiplier*					Avg.	Variability
	Mean	+2 SD	+1 SD	-1 SD	-2 SD		
0	5.09	5.06	5.14	5.14	5.22	5.13	±0.08
1	3.26	3.25	3.30	3.27	3.28	3.27	±0.03
2	2.60	2.57	2.62	2.62	2.64	2.61	±0.04
3	2.24	2.21	2.26	2.26	2.28	2.25	±0.03
4	2.00	1.96	2.01	2.02	2.04	2.00	±0.04
5	1.82	1.79	1.83	1.84	1.86	1.83	±0.03
6	1.68	1.64	1.69	1.70	1.73	1.69	±0.04
7	1.56	1.52	1.56	1.58	1.61	1.57	±0.05
8	1.46	1.43	1.46	1.49	1.51	1.47	±0.04
9	1.37	1.34	1.38	1.40	1.42	1.38	±0.04
10	1.30	1.27	1.30	1.32	1.35	1.31	±0.04
11	1.24	1.20	1.24	1.26	1.29	1.24	±0.05
12	1.18	1.14	1.17	1.20	1.23	1.18	±0.05
13	1.12	1.07	1.11	1.15	1.18	1.13	±0.06
14	1.07	1.03	1.06	1.09	1.12	1.08	±0.03
15	1.03	1.01	1.04	1.05	1.06	1.04	±0.03
16	1.01	1.00	1.02	1.02	1.03	1.02	±0.02
17	1.00	1.00	1.02	1.01	1.01	1.01	±0.01
18	1.00	1.00	1.01	1.00	1.00	1.00	±0.00

*SD = standard deviation.

TABLE IV
TIBIAL MULTIPLIERS FOR GIRLS BY PERCENTILE GROUP

Age (yrs.)	Multiplier*					Avg.	Variability
	Mean	+2 SD	+1 SD	-1 SD	-2 SD		
0	4.76	4.58	4.54	4.58	4.67	4.63	±0.11
1	2.99	3.03	3.01	2.98	2.95	2.99	±0.04
2	2.39	2.44	2.41	2.36	2.33	2.39	±0.06
3	2.06	2.10	2.08	2.04	2.02	2.06	±0.04
4	1.84	1.84	1.84	1.83	1.83	1.84	±0.01
5	1.67	1.67	1.67	1.67	1.67	1.67	±0.00
6	1.54	1.53	1.53	1.54	1.55	1.54	±0.01
7	1.43	1.42	1.42	1.44	1.45	1.43	±0.02
8	1.34	1.32	1.33	1.35	1.36	1.34	±0.02
9	1.26	1.24	1.25	1.27	1.29	1.26	±0.03
10	1.18	1.16	1.17	1.20	1.22	1.19	±0.03
11	1.12	1.09	1.10	1.14	1.16	1.12	±0.04
12	1.06	1.04	1.05	1.08	1.09	1.06	±0.03
13	1.02	1.01	1.02	1.03	1.04	1.03	±0.02
14	1.00	1.00	1.00	1.00	1.00	1.00	±0.00
15	1.00	1.00	1.00	1.00	1.00	1.00	±0.00
16	1.00	1.00	1.00	1.00	1.00	1.00	±0.00

*SD = standard deviation.

Abbildung 6: Tabellen altersspezifischer Multiplikatoren: Femorale Multiplikatoren für Knaben (links) und tibiale Multiplikatoren für Mädchen (rechts)(entnommen aus (85) mit Erlaubnis von Wolters Kluwer Health, Inc.).

Es stellte sich heraus, dass die Werte der verschiedenen Perzentilen kaum voneinander abweichen. Die Multiplikatoren sind daher unabhängig von der Perzentile und eine Bestimmung derselben muss nicht erfolgen. Wohl aber finden sich Unterschiede zwischen den Geschlechtern und Altersgruppen. Die Multiplikatoren von Femur und Tibia sind einander so ähnlich, dass man sie in einen einzelnen Multiplikator für die gesamte untere Extremität zusammenfassen kann. Es ist auch möglich, das verbleibende Wachstum einzelner Wachstumsfugen zu berechnen. Stark vereinfacht lässt sich sagen, dass 71% des femoralen Längenwachstums auf den distalen Femur zurückzuführen ist. Die proximale Tibia ist für 57% der gesamten Tibialänge verantwortlich. Multipliziert man also das verbleibende Längenwachstum des gesamten Femur mit 0,71, erhält man das Wachstum, das von der distalen Wachstumsfuge noch zu erwarten ist. Diese Berechnung lässt sich analog für die Tibia durchführen und ist besonders in der Planung einer Epiphysiodese von Interesse. Paley et al. entwickelten noch weitere Formeln zur Voraussage des Wachstums für unterschiedliche Fragestellungen rund um das Thema Beinlängendifferenz und die damit verbundenen chirurgischen Eingriffe (13).

Die Multiplier Method wurde mit 21 anderen, teils sehr unterschiedlichen Datensammlungen verglichen. Es zeigten sich nur geringe Abweichungen der Ergebnisse, weshalb die Autoren schlussfolgerten, dass die Multiplikatoren unabhängig von Faktoren wie ethnischer Herkunft, Nationalität, Generation und sozioökonomischem Status sind (13). Die Multiplier Method ist schnell, einfach und in jedem Alter anwendbar, was ein Vorteil gegenüber dem Moseley Straight-Line Graph und anderen auf den Daten von Anderson et al. basierenden Methoden ist. Wie viele andere Methoden wird auch die Multiplier Method präziser, je älter das untersuchte Kind ist. Besonders praktisch ist, dass die für die Methode notwendigen Unterlagen handlich und dadurch schnell verfügbar sind. Die Multiplier Method verwendet das CA, weshalb Abweichungen des SA vom CA durch die Methode nicht berücksichtigt werden (13). Es existieren auch Multiplikatoren für die obere Extremität (86).

1.4 Grundlagen des maschinellen Lernens und Deep Learning

Maschinelles Lernen bezeichnet die Fähigkeit einer künstlichen Intelligenz (KI, englisch Artificial Intelligence, AI) sich Wissen anzueignen, indem sie Muster in Datenmengen erkennt (87). Informationen über Eigenschaften eines Objektes, die man der KI liefert, werden als Feature bezeichnet (88). Ein Neural Network imitiert die Fähigkeit des menschlichen Gehirns, aus Beobachtungen bzw. Erfahrungen zu lernen. Ein Deep Learning Neural Network ist, vereinfacht ausgedrückt, ein vielschichtiges Neural Network (87). Um zu verstehen, wie ein komplexes Neural Network aufgebaut ist, müssen zunächst die Grundlagen betrachtet werden.

1.4.1 Erklärung der Arbeitsweise

Beispielsweise sollen auf einem Fließband viele Objekte zweier Arten automatisch voneinander unterschieden werden (88). Dazu soll ein Programm einzelne Objekte den zwei Arten zuordnen. Um das zu bewerkstelligen werden mehrere Arbeitsschritte durchlaufen: Im ersten Schritt, dem Preprocessing, werden die Bilder der einzelnen Objekte auf eine Weise bearbeitet und vereinfacht, sodass das Objekt jeweils einzeln zu erkennen ist. Jegliche überflüssigen Informationen sollten von dem Bild entfernt werden, da diese für die Erkennung des Objektes irrelevant sind. Der für die Auswertung relevante Anteil des Bildes wird auch als Region of Interest (ROI) bezeichnet. Im nächsten Schritt erfolgt die Feature Extraction. Dabei wird die Information auf den Bildern auf ein oder mehrere Features reduziert, beispielsweise die Dimensionen oder Farbe des Objektes. Diese Features werden von dem Programm gemessen. Im letzten Schritt, der Classification, trifft die Software anhand der gemessenen Werte die Entscheidung, welcher der beiden Kategorien das Objekt angehört. Dazu muss natürlich das Wissen vorhanden sein, welche Features in den Objekten wie ausgeprägt sind. Die Wahl der Features und deren Anzahl sind ausschlaggebend für das Ergebnis der Classification. Es müssen ausreichend Features verwendet werden, um das Problem adäquat zu beschreiben. Werden zu viele Features definiert, kann die Software zwar die Objekte des Trainingsdatensatzes fehlerlos zuordnen, jedoch ist eine Generalisierung, also die Anwendung auf neue und möglicherweise leicht abweichende Datenmengen, nicht mehr möglich (88). Wichtig ist auch, dass die gemessenen Features des Trainingsdatensatzes repräsentativ für die Objekte sind (87).

1.4.2 Aufbau eines Neural Network

Ein Neural Network vereint Feature Extraction und Classification in einem Schritt (87). Im Gegensatz zu einfacheren Algorithmen, denen Features und deren Gewichtung vorgegeben werden, baut sich ein Neural Network eine eigene Hierarchie auf. Diese besteht aus mehreren Ebenen, genannt Layers, die miteinander in Verbindung stehen. Das erste Layer einer Bilderkennungssoftware erkennt beispielsweise Kanten und Farbübergänge. Das nächste Layer fügt diese zu komplexeren Strukturen zusammen. So entstehen immer abstraktere Layers, die auf den vorhergehenden, simpleren Konzepten basieren. Je nach Komplexität der Software kann die Anzahl der Layers stark variieren. Ein Neural Network mit vielen übereinander liegenden Ebenen wird als „tief“ bezeichnet – daher stammt der Ausdruck Deep Learning. Stellt man diese Architektur bildlich dar, erhält man eine Art Entscheidungsbaum, der Graph genannt wird. Er besteht aus Knoten, die Entscheidungen repräsentieren. Das Ergebnis eines Knotens wird an den nächsten weitergegeben, wodurch eine Hierarchie entsteht (87).

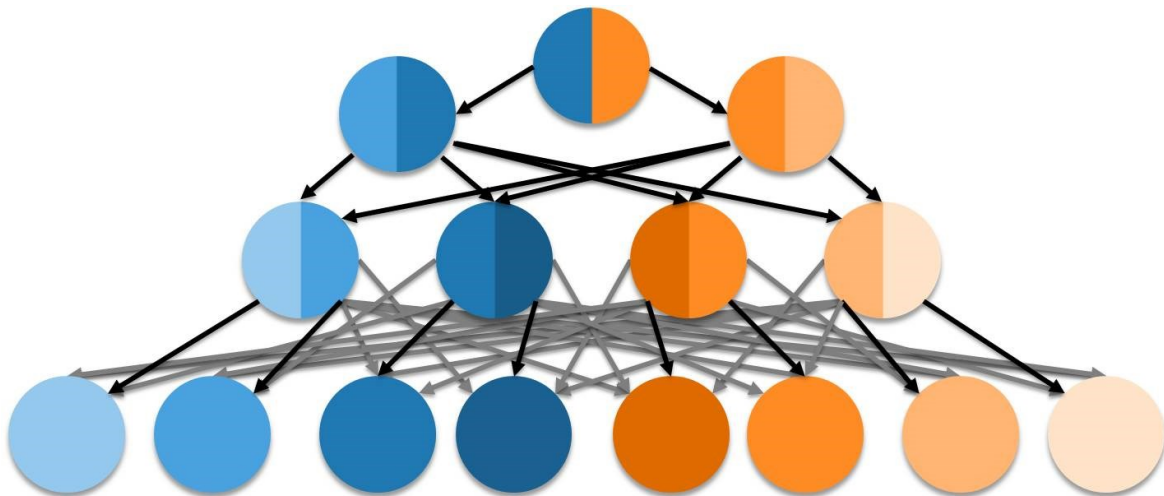


Abbildung 7: Schema eines klassischen Neural Network: An den Knoten werden Entscheidungen getroffen, die an die folgenden Ebenen weitergegeben werden. Die Knoten einer Schicht sind mit allen Knoten der nächsten Schicht verbunden. Die Gewichte entlang dieser Verbindungen sind die Parameter, die im Training der Software adaptiert werden (eigene Darstellung).

1.4.3 Training und Validierung

Zunächst wird die Software mithilfe eines Datensatzes, in welchem alle Ergebnisse bereits bekannt sind, trainiert (87). Die erste Schätzung der Software wird mit hoher Wahrscheinlichkeit falsch ausfallen. Danach korrigiert sie anhand der korrekten Werte,

die ihr im Anschluss zugeführt werden, ihre Parameter und führt Anpassungen durch. Diese laufen in die entgegengesetzte Richtung der anfänglichen Berechnungen, weshalb sie auch Backpropagation genannt werden. Die einzelnen Gewichtungen der Entscheidungen bzw. Knoten werden so korrigiert, dass der Wert am Ende des Graphen stimmt. Dieser Vorgang wird viele Male wiederholt. Im erwünschten Fall findet die Software schließlich jene Features, die mit dem Ergebnis korrelieren und setzt deren Gewichtung in ihrer Entscheidungsfindung entsprechend hoch an. Ein hohes Maß an irrelevanten und damit störenden Informationen in den Daten erschwert diesen Vorgang. Darum ist es wichtig, im Preprocessing ein größtmögliches Ausmaß an irrelevanten Informationen zu entfernen (87).

In der Validierung wird die Software mit einem neuen, ihr unbekanntem Datensatz konfrontiert. Dann zeigt sich, wie gut relevante Information erkannt werden konnte und ob mit der Architektur des Neural Networks eine Generalisierung möglich ist (87).

1.4.4 Probleme und Lösungsstrategien

Abgesehen davon, dass die Qualität der Ergebnisse eines Neural Network stark durch den präsentierten Datensatz und dessen Repräsentativität bestimmt ist, ergeben sich zwei weitere Probleme: Underfitting und Overfitting (87). Underfitting bedeutet, dass im Training keine Reduktion der Fehler stattfindet. Beim Overfitting werden die Fehler im Training zwar reduziert, in der Validierung jedoch nicht. So kann es passieren, dass sich eine Software nach zahlreichen Trainingsdurchläufen dem Datensatz so gut angepasst hat, dass ihre Ergebnisse beinahe fehlerfrei sind. Allerdings fehlt ihr dadurch auch die Fähigkeit der Generalisierung, sie ist also auf unbekannte Daten nicht anwendbar. Die Abweichungen werden in der Validierung von Durchlauf zu Durchlauf nicht geringer (87). Je größer der verfügbare Datensatz ist, desto eher kann Overfitting vermieden werden. Augmentation ist eine Strategie, um dieses Risiko zu reduzieren (73). Dabei werden die Datensätze durch geringe, oft zufällige Veränderungen der Bilder künstlich erweitert. Zum Beispiel können Rotationen vorgenommen, Helligkeit verändert oder Noise („Rauschen“) eingefügt werden. Dadurch soll eine bessere Resistenz des Neural Network gegenüber derartiger Störfaktoren erzielt werden (73).

Eine sehr hohe Anzahl an Layers in Deep Learning Neural Networks kann dazu führen, dass bei der Backpropagation die Fehlerkorrektur nicht mehr bis in die obersten Layers durchgeführt werden kann, da das Ausmaß der Anpassung von Layer zu Layer abnimmt. Eine effektive Korrektur ist dann nicht mehr möglich. Residual Networks bieten eine Lösung dieses Problems.(89) Anders als herkömmliche Plain Networks weisen sie zusätzliche Verbindungen zwischen nicht benachbarten Ebenen auf.

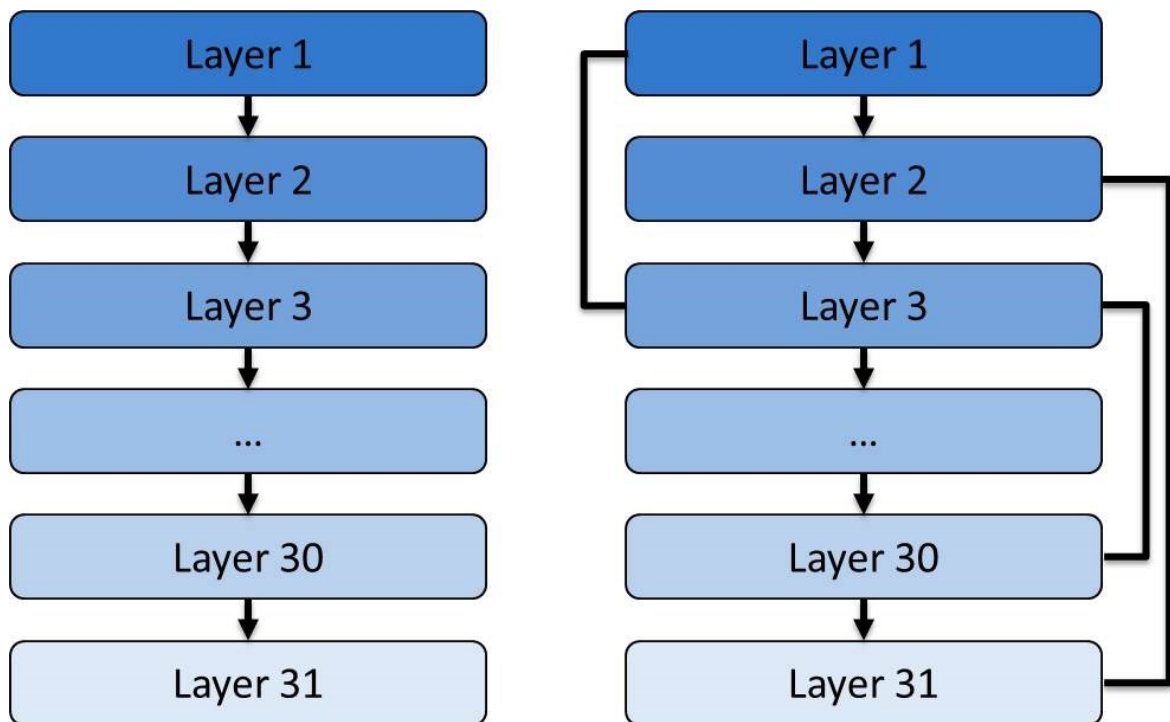


Abbildung 8: Schematische Darstellung unterschiedlicher Neural Networks: Plain Network (links) und Residual Network (rechts). Das Residual Network weist zusätzliche Verbindungen zwischen nicht benachbarten Ebenen auf (eigene Darstellung).

Auf die Bedeutung des Preprocessings wurde bereits eingegangen. Im Falle dreidimensionaler MRT-Aufnahmen bedeutet das, dass die ROI (bzw. das Volume of Interest) nicht in allen Aufnahmen gleich positioniert ist, da Patientinnen/Patienten nicht stets gleich gelagert werden können und auch Unterschiede hinsichtlich der Anatomie vorliegen. Sowohl Translationen als auch Rotationen sind möglich. Die Anordnung verschiedener Bilder in möglichst gleicher Ausrichtung wird Registrierung genannt (90). So können beispielsweise Röntgen- und CT-Aufnahmen oder Bilder von Verlaufskontrollen übereinander gelegt werden, um eine bessere Beurteilung zu ermöglichen (90).

1.5 Hypothese, Fragestellung und Zielsetzung

Es ist anzunehmen, dass keine der weitläufig verwendeten Methoden unkritisch auf Kinder der österreichischen Durchschnittsbevölkerung anwendbar ist, besonders wenn es um die Voraussage der Erwachsenengröße geht. Die Handskelettalterbestimmung nach Greulich und Pyle zeigte in der Studie von Gröll et al. zwar eine starke Korrelation zwischen geschätztem SA und CA, jedoch hat die Methode Limitationen, die bereits erläutert wurden (24). Die Aufnahme eines Röntgenbildes stellt immer auch eine Strahlenbelastung für die Betroffene/den Betroffenen dar und die für die Standards nach Greulich und Pyle verwendete Studienpopulation gilt als veraltet. Ungenauigkeiten bei der Bestimmung des SA verschlechtern die Präzision der Größenvoraussage, was wiederum therapeutische Konsequenzen haben kann (13). Ein Gewinn an Genauigkeit ist deshalb auch ein Gewinn an Qualität der Therapie für Patientinnen/Patienten.

Wie beschrieben erfolgt die Schätzung des SA und somit auch der Erwachsenengröße sowie der Dauer bis zur Erreichung derselben bisher zum überwiegenden Anteil auf Basis röntgenologisch gewonnener Bildinformation. In den letzten Jahren wurde vermehrt die MRT als strahlungsfreie, alternative radiologische Bildgebungsmodalität zur SA-Schätzung herangezogen (55-57, 62). Die Wahl der MRT als bildgebendes Verfahren zur Abschätzung des SA hat mehrere Gründe. Die MRT verspricht zusätzlich zum Vorteil der wegfallenden Strahlenbelastung eine höhere Genauigkeit der Altersschätzung aufgrund des möglichen Weichgewebskontrasts, durch den knorpelige Strukturen besser beurteilbar sind (55). Im Gegensatz zum Röntgen kann die MRT dreidimensionale Informationen von den Wachstumsfugen liefern. Dieser zusätzliche Informationsgewinn ist auch für die orthopädische Operationsplanung von Interesse. Die Aufnahme dauert zwar länger als bei einem Röntgenbild und ist deutlich kostenintensiver, jedoch liegt bei vielen orthopädischen Patientinnen/Patienten aufgrund vorangegangener diagnostischer Fragestellungen bereits eine aktuelle MRT-Untersuchung vor, sodass in diesen Fällen kein zusätzlicher zeitlicher und finanzieller Mehraufwand entstehen würde. Das Wegfallen der Strahlenbelastung begünstigt außerdem regelmäßige Verlaufskontrollen, die die Präzision der Größenvorhersage erhöhen. Eine MRT-Datensammlung würde in diesem Sinne helfen, eine aktuelle, modernisierte Datenbasis für die SA-Schätzung von Kindern und Jugendlichen aufzubauen. Sowohl aus ethischen als auch aus rechtlichen Gründen ist

diese Art der röntgenbasierten Datensammlung hinsichtlich der anzuwendenden Strahlenhygiene heute nicht mehr möglich.

Abgesehen davon, dass nicht nur bei Beinlängenunterschieden und Achsabweichungen in vielen Fällen eine MRT des Kniegelenks bereits vorhanden ist oder angestrebt wird, bietet das Kniegelenk weitere Vorteile für die Bestimmung des SA. Die Wachstumsfugen des Kniegelenks, respektive des distalen Femurs sowie der proximalen Tibia, bieten sich für eine Untersuchung an, da auf diese beiden Wachstumsfugen ein überwiegender Anteil des Längenwachstums der unteren Extremitäten zurückgeführt werden kann (91). Für die Voraussage der Erwachsenengröße erscheint dies sinnvoll. Dennoch sind Studien, die das SA mittels MRT des Kniegelenks erheben, selten (55-59).

Eine Software aus dem Deep Learning-Bereich hingegen, die vor allem Objektivität und Zeitersparnis als Vorteile aufweist, böte sich in diesem Sinne für die Fragestellung nach der SA-Bestimmung oder in weiterer Folge auch für die Voraussage der Erwachsenengröße an. Auf Knie-MRT-Daten angewandt würde auf diese Art die Möglichkeit bestehen, das SA aus standardisierten MRT-Aufnahmen des wachsenden Knies mittels darauf trainierter selbstlernender Software berechnen zu lassen. Somit könnte ein neuer Weg in der Bestimmung des SA aufgezeigt werden.

Die Vorteile der MRT und die bereits beschriebenen Entwicklungen im Bereich selbstlernender Softwares sind Beweggründe, die Pilotstudie „MRT-basierte Knochenwachstumsschätzung zur Planung von Eingriffen in der Kinderorthopädie“ durchzuführen. Es stellt sich die Frage, ob ein solches Software-basiertes Modell die Bestimmung des SA aus den gängigen standardisierten MRT-Daten des wachsenden Kniegelenks tatsächlich ermöglicht, und diese Art der Altersschätzung auch in ausreichender Genauigkeit erfolgen kann. Ziel dieser Pilotstudie ist die Kalibrierung einer Deep-Learning-Software auf Basis retrospektiver Daten, um ein aussagekräftiges Modell zur SA-Bestimmung an MRT-Daten des Kniegelenks im Wachstumsalter auszutesten, und um anhand der Ergebnisse einen Ausblick auf etwaige zukünftige Anwendungen wie die automatisierte Abschätzung des verbleibenden Restwachstums wagen zu können.

2 Material und Methoden

Die vorliegende Studie wurde an der Abteilung für Kinderorthopädie der Medizinischen Universität Graz in Kooperation mit der Abteilung für Kinderradiologie und dem Ludwig Boltzmann Institut für Klinisch-Forensische Bildgebung durchgeführt.

An der Abteilung für Kinderorthopädie der Medizinischen Universität Graz sind alle kinderorthopädischen Subspezifitäten vertreten. Das Zentrum behandelt Kinder und Jugendliche bis zum Alter von 18 Jahren stationär und ambulant. Patientinnen/Patienten mit speziellen Krankheitsbildern wie z.B. neuroorthopädischen Erkrankungen werden oft auch über dieses Alter hinaus betreut. Das Einzugsgebiet umfasst die Steiermark sowie angrenzende Bundesländer und Nachbarstaaten (z.B. Slowenien).

Das Ludwig Boltzmann Institut für Klinisch-Forensische Bildgebung (LBI CFI) hat seinen Forschungsschwerpunkt im Einsatz bildgebender Verfahren in der klinischen Rechtsmedizin. Das interdisziplinäre Forschungsteam erarbeitet die Grundlagen für eine klinisch-forensische Anwendung der MRT- und CT-Bildgebung, um die Qualität forensischer Gutachten zu verbessern und somit zu einer Erhöhung der Rechtssicherheit beizutragen. Auch die nicht-invasive Altersschätzung auf Basis von MRT zählt zu den Forschungsgebieten des Instituts. Gemeinsam mit der Medizinischen Universität Graz betreibt das LBI CFI die Klinisch-Forensische Untersuchungsstelle, eine Anlaufstelle für Menschen, die von körperlicher und/oder sexueller Gewalt, Kindesmissbrauch oder Kindesmisshandlung betroffen sind.

Die zuständige Ethikkommission in Graz erteilte die Zustimmung zur Durchführung der Studie (EK-Nummer: 1396/2016). Da es sich um eine retrospektive Studie handelt, war das Einholen von Einwilligungserklärungen der Studienteilnehmerinnen/Studienteilnehmer nicht notwendig. Durch die Anonymisierung aller Daten wurde sichergestellt, dass keinerlei sensitive Patienteninformation an die Mitarbeiterinnen/Mitarbeiter der Studie am LBI CFI weitergegeben wurde.

2.1 Studiendesign und -ablauf

Es handelt sich um eine monozentrische Pilotstudie mit einer retrospektiven Datensammlung und -auswertung. Zunächst wurde ein Datensatz aller Patientinnen/Patienten unter 18 Jahren mit zwei oder mehr MRT-Untersuchungen des

Kniegelenks derselben Körperseite, die im Zeitraum 2000-2016 an der Abteilung für Radiologie der Medizinischen Universität Graz durchgeführt worden waren, erhoben und auf Ausschlusskriterien überprüft. Die Datensammlung inklusive der Sammlung der Datensätze der MRT-Untersuchungen wurde an der Abteilung für Kinderorthopädie der Universitätsklinik für Kinder- und Jugendchirurgie angelegt. Nach Anonymisierung wurden diese an das LBI CFI übermittelt. Dort erfolgten Softwareentwicklung und -validierung zur Bestimmung des SA durch die Software. In Kooperation mit dem LBI CFI wurden die Ergebnisse ausgearbeitet.

2.1.1 Zielgrößen

Mit Hilfe der erstellten Datensammlung wurde die Hypothese überprüft, dass eine objektive Wachstumlängenschätzungssoftware auf Basis von Algorithmen des maschinellen Lernens und der medizinischen Bildanalyse eine Schätzung des SA anhand von MRT-Daten des Knies liefern kann. Die Differenz zwischen errechnetem und tatsächlichem Alter zeigt, wie genau die Software arbeitet und ob diese Methode in Zukunft sinnvoll anwendbar sein könnte. Doch nicht nur die Abweichung der Schätzung ist ein Maß für die Qualität der Methode. Auch die Lernfähigkeit, also die Abnahme der Fehler mit jedem neuen Versuch, spielt eine Rolle und wurde durch das LBI CFI evaluiert. Die Zielgrößen der Datensammlung (z.B. CA, Geschlecht und MRT-Aufnahmen) werden im folgenden Kapitel genau erläutert.

2.2 Datenakquirierung

Zunächst wurden alle Patientinnen/Patienten im Alter von 0 bis 18 Jahren, die im Zeitraum 2000-2016 an der Abteilung für Radiologie des Universitätsklinikums Graz mehrere MRT-Untersuchungen des Kniegelenks erhielten, identifiziert und eine Liste angelegt. Dies erfolgte EDV-gestützt mittels des Dokumentationssystems MEDOCS und des Bilddokumentations- und -betrachtungsprogrammes PACS. MEDOCS ist ein SAP (System Applications and Products in Data Processing, Walldorf, Germany)-basiertes System, das an der gesamten Universitätsklinik Graz in Verwendung ist. Mit dem Bildarchivierungs- und Kommunikationssystem PACS (Picture Archiving and Communication System) können an der Universitätsklinik Graz medizinische Bilddaten unterschiedlicher Modalitäten (z.B. Röntgen, CT, MRT) gespeichert und aufgerufen werden. Die Dateien werden im DICOM-Format angelegt. Die Patientenliste mit Patientennamen, Fall- und Patientennummern sowie jeweiligem Geburtsdatum wurde in eine Microsoft Excel-Tabelle importiert. Die zur Vervollständigung der Excel-Tabelle notwendigen Daten jeder Patientin/jedes Patienten wurden einzeln aus MEDOCS recherchiert und in die Tabelle eingetragen.

2.2.1 Ein- und Ausschlusskriterien

Die initial erhobene Liste enthielt die oben beschriebene Patientengruppe unabhängig davon, welche Seite untersucht wurde. Daher wurden im ersten Schritt jene Patientinnen/Patienten ermittelt, bei denen mindestens zwei MRT-Aufnahmen des ipsilateralen Kniegelenks verfügbar waren. Sämtliche anderen Patientinnen/Patienten wurden aus der Studie ausgeschlossen, da aus einer einzigen Untersuchung keine longitudinale Information gewonnen werden kann. Lagen von einer Patientin/einem Patienten beispielsweise zwei MRT-Aufnahmen der linken Seite und eine der rechten Seite vor, wurden nur die der linken Seite in die Studie aufgenommen. Da innerhalb kurzer Zeit mit keinen oder nur geringen Veränderungen der kindlichen Wachstumsfugen zu rechnen ist, entschied man sich für einen Mindestabstand von sechs Monaten zwischen zwei MRT-Untersuchungen. In wenigen Fällen wurde der Mindestabstand um einige Tage unterschritten, diese Patientinnen/Patienten wurden aufgrund der geringen

Abweichung eingeschlossen. Patientinnen/Patienten, deren zwei Untersuchungen in einem geringeren zeitlichen Abstand durchgeführt wurden, wurden ausgeschlossen.

Um davon ausgehen zu können, dass die Studienpopulation einen repräsentativen Querschnitt der im Wachstum befindlichen Population Österreichs darstellt und als solcher eine Normalverteilung bezüglich des SA aufweist, mussten weitere Ausschlusskriterien formuliert werden. Ausgeschlossen wurden Patientinnen/Patienten, die zum Zeitpunkt der MRT-Untersuchungen an einer Grunderkrankung litten, die das Wachstum erwiesenermaßen beeinträchtigt. Nur so konnte das CA statt des SA verwendet werden. Ebenso wurden Patientinnen/Patienten, die eine posttraumatisch oder durch Osteomyelitis oder Osteoarthritis veränderte Gelenks- bzw. Knochenmorphologie aufwiesen, von der Studie ausgeschlossen. Weitere Ausschlusskriterien waren Leukämien und Tumorerkrankungen unter Chemotherapie sowie komplexe knöcherne oder weichteilige Veränderungen des Kniegelenks. Die Entscheidung über den Ausschluss einer Patientin/eines Patienten erfolgte individuell und fallbezogen durch Einsichtnahme in die Krankengeschichte in MEDOCS sowie gegebenenfalls der MRT-Untersuchung.

Im nächsten Schritt wurden die Wachstumsfugen der verbliebenen MRT-Aufnahmen im PACS untersucht. Wies der Knochen kein „Knorpelsignal“ und keine Veränderung zum Vorbefund an dieser Stelle auf, wurde die Wachstumsfuge als geschlossen bezeichnet. Lag ein Zustand nach Schluss der Wachstumsfugen vor, wurden alle Untersuchungen nach derjenigen, in der der Endzustand erstmals sichtbar wurde, verworfen. In jedem Fall erfolgte die Inspektion der später angefertigten MRT-Aufnahmen auf starke Veränderungen im Bereich der ehemaligen Wachstumsfugen. Somit liegt bei allen Patientinnen/Patienten maximal eine Aufnahme bei vollendetem Wachstum vor. Bei nicht eindeutigem Befund wurde der Datensatz nicht ausgeschlossen. Waren die Wachstumsfugen in allen Untersuchungen einer Patientin/eines Patienten eindeutig geschlossen, wurde diese/dieser ausgeschlossen, da zwischen den beiden Untersuchungszeitpunkten keine für diese Studie relevante Veränderung eingetreten sein kann.

2.2.2 Datensammlung

Folgende Zielgrößen wurden erhoben und in der Excel-Tabelle erfasst:

- Vor- und Nachname
- Geschlecht
- Geburtsdatum
- Ethnische Herkunft
- Datum der MRT-Untersuchung
- Alter der Patientin/des Patienten zum Zeitpunkt der Untersuchung
- Untersuchungsseite (links/rechts)
- Nummerierung der MRT-Untersuchung des einzelnen Knies
- Zeitlicher Abstand zur letzten MRT-Untersuchung desselben Knies
- Zeitlicher Abstand zur ersten MRT-Untersuchung desselben Knies
- Größe und Gewicht
- Kniespezifische Diagnose bzw. Fragestellung der MRT-Untersuchung
- Voroperationen am selben Knie mit Datum der Operation
- Andere relevante Erkrankungen
- Datum einer oder mehrerer Röntgenuntersuchungen desselben Knies (innerhalb von sechs Monaten zur MRT-Untersuchung)
- Datum einer oder mehrerer Röntgenuntersuchungen der linken Hand und das daraus ermittelte SA (innerhalb von sechs Monaten zur MRT-Untersuchung)
- Datum einer oder mehrerer CT-Untersuchungen desselben Knies (innerhalb von sechs Monaten zur MRT-Untersuchung)

Weiters wurden drei Nummerierungen geführt: Eine für die Patientinnen/Patienten, eine für die einzelnen Kniegelenke und eine für die MRT-Untersuchungen.

Alle MRT-Untersuchungen dieser Studie wurden an der Abteilung für Kinderradiologie des LKH-Universitätsklinikums Graz durchgeführt. Die MRT-Bilddaten stammen aus dem 1,5 Tesla Gerät „Symphony TIM“ (Siemens Healthineers, Erlangen, Deutschland) mit „Dedicated knee coils“. Die Untersuchungen erfolgten in Rückenlage. Abhängig von der klinischen Fragestellung kamen unterschiedliche Sequenzen zum Einsatz. Ein typisches Untersuchungsprotokoll bestand aus folgenden Sequenzen:

- Axiale, T2-gewichtete Turbo-Spin-Echo-Sequenz mit Fettsättigung
- Sagittale, T2-gewichtete Turbo-Spin-Echo-Sequenz
- Sagittale, T2-gewichtete Gradienten-Echo-Sequenz (Double Echo Steady State)
- Koronare, T2-gewichtete Turbo-Inversion-Recovery-Magnitude-Sequenz
- Koronare, T1-gewichtete Spin-Echo-Sequenz

Die MRT- und Röntgenaufnahmen der Kniegelenke wurden aus PACS exportiert und gemeinsam mit der Excel-Tabelle auf einem externen Datenträger gespeichert. Es erfolgte die Anonymisierung aller Daten und die Übergabe an das LBI CFI. In dieser Studie wurden nur MRT-Bilddaten in sagittaler Schnittebene und T2-gewichteter Sequenz ausgewertet, da diese Ebene in den meisten Fällen verfügbar war.

2.3 Altersschätzung durch Neural Network

An den MRT-Aufnahmen im DICOM-Format erfolgte am LBI CFI das Preprocessing. Für die Registrierung wurde die Software Elastix, die aus mehreren Algorithmen der medizinischen Bildregistrierung besteht, verwendet (90). Die Region of Interest (ROI) bzw. das dreidimensionale Volume of Interest wurde manuell auf einer MRT-Aufnahme ausgewählt. Alle weiteren Aufnahmen wurden auf diese registriert und dementsprechend transformiert und zugeschnitten.

Die Aufteilung der verfügbaren registrierten Datensätze auf Trainings- und Validierungsset erfolgte randomisiert. Die Altersverteilung innerhalb beider Sets wurde möglichst gleichmäßig gehalten. Eine Augmentation der MRT-Aufnahmen wurde durchgeführt. Es wurden zufällige Skalierungen, Rotationen, Translationen und Intensitätsshifts sowie Noise eingefügt.

Es erfolgten mehrere Versuche (Experimente) mit Deep Learning Neural Networks unterschiedlicher Architektur. Es kamen auch modifizierte Residual Networks mit 42 und 50 Layers (ResNet42, ResNet50) zum Einsatz.

3 Ergebnisse – Resultate

3.1 Ergebnisse der Datenakquirierung

Die Suche in MEDOCS und PACS nach Patientinnen/Patienten im Alter von 0 bis 18 Jahren, die im Zeitraum 2000-2016 an der Abteilung für Radiologie des Universitätsklinikums Graz mehrere MRT-Untersuchungen des Knies erhielten, ergab 592 MRT-Aufnahmen von 233 Patientinnen/Patienten. Nach Ausschluss einiger Patientinnen/Patienten bzw. MRT-Aufnahmen nach den oben genannten Kriterien bestand die Datensammlung aus 317 MRT-Aufnahmen von 121 Patientinnen/Patienten. Häufigster Ausschlussgrund war zu kurzer zeitlicher Abstand zwischen Untersuchungen, seltener bereits geschlossene Wachstumsfugen oder schwere Grunderkrankungen. Die Zusammensetzung der Datensammlung wird anhand der folgenden Abbildungen erläutert.

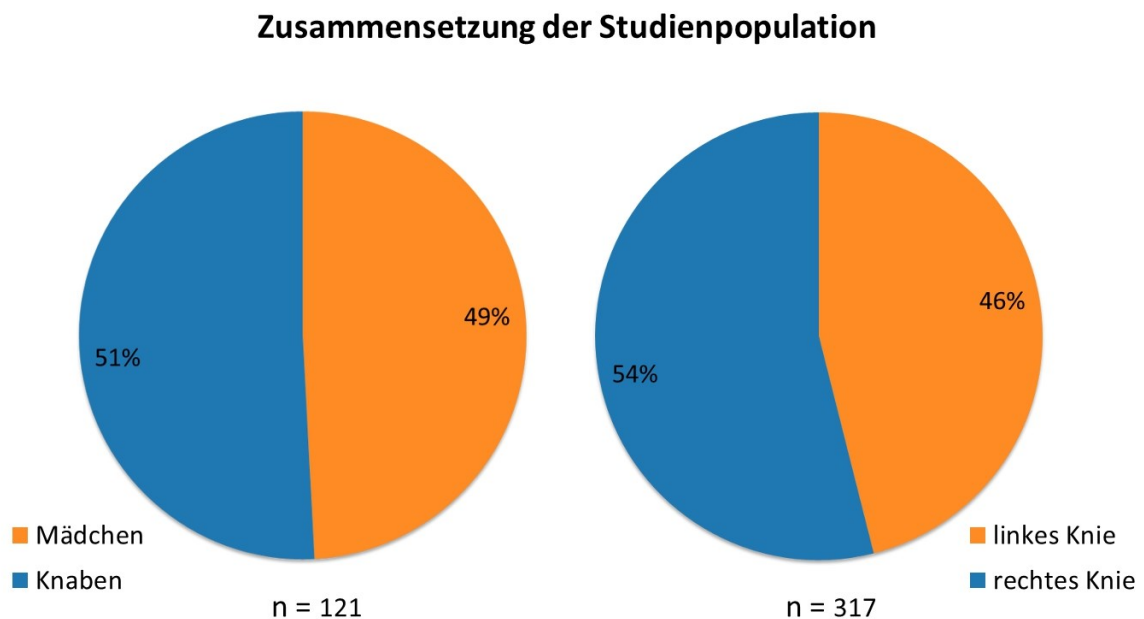


Abbildung 9: Zusammensetzung der Studienpopulation: Unter 121 Patientinnen/Patienten waren 59 Mädchen und 61 Knaben. Von 317 MRT-Aufnahmen zeigten 146 das linke und 171 das rechte Knie (eigene Darstellung).

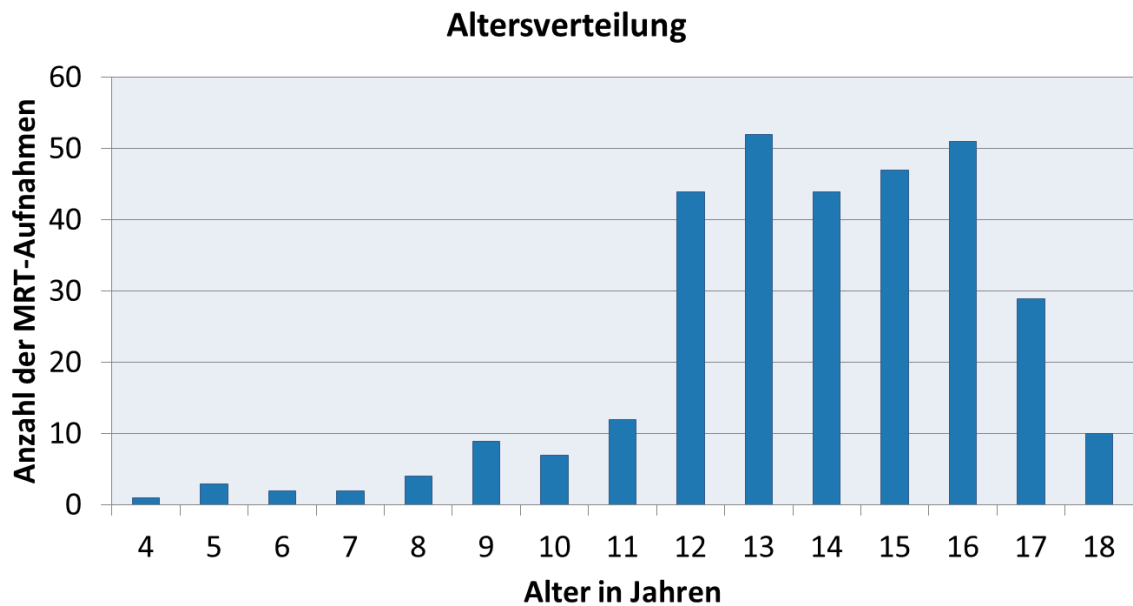


Abbildung 10: Altersverteilung der Studienpopulation: Von den 317 MRT-Untersuchungen wurde ein großer Anteil im Alter von 12 bis 16 Jahren durchgeführt. Besonders jüngere Lebensalter sind unterrepräsentiert. Das durchschnittliche Alter bei der Untersuchung lag bei 14,24 Jahren (eigene Darstellung).

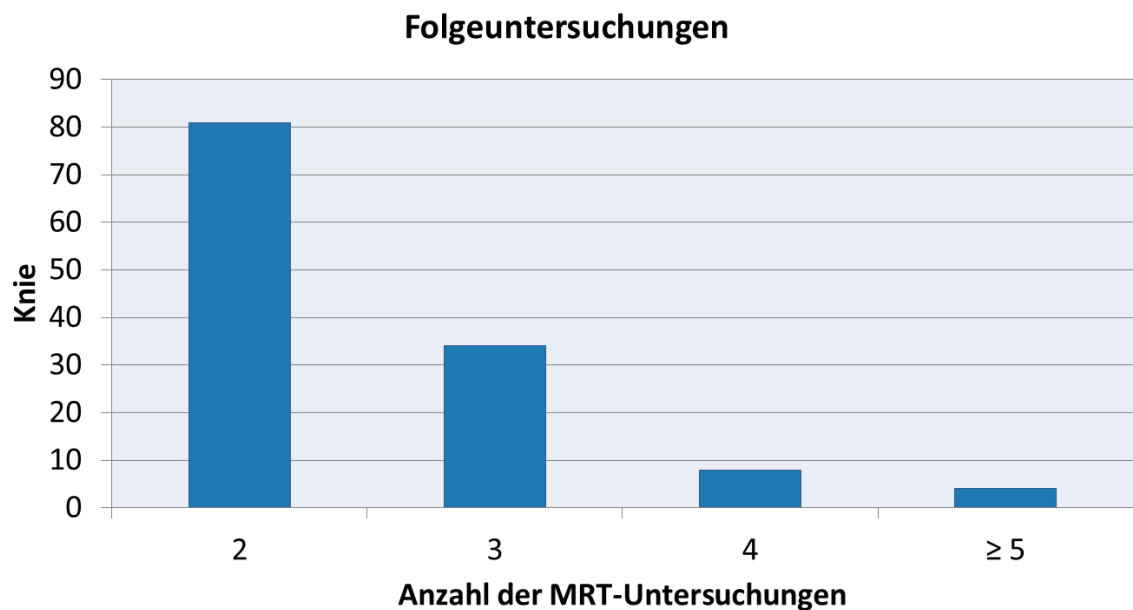


Abbildung 11: Anzahl der MRT-Untersuchungen pro Knie: Da bei mehreren Patientinnen/Patienten beide Kniegelenke untersucht wurden, werden an dieser Stelle die Untersuchungen pro Knie angegeben. In den meisten Fällen lagen nur zwei MRT-Aufnahmen vor. Nur in vier Fällen waren fünf oder mehr Untersuchungen durchgeführt worden. Für die SA-Bestimmung wurde allerdings nur eine MRT-Aufnahme pro Patientin/Patient verwendet (eigene Darstellung).

Die Indikationen für MRT-Untersuchungen der Studienpopulation waren in vielen Fällen traumatisch bedingt. Patellaluxationen, Rupturen des vorderen Kreuzbandes und Meniskusverletzungen waren häufige Diagnosen. Ebenso konnte eine hohe Anzahl an Fällen von Osteochondrosis dissecans beobachtet werden. Bei 179 von 317 MRT-Aufnahmen bestand ein Zustand nach chirurgischem Eingriff. Einige Patientinnen/Patienten wiesen in Verlaufsuntersuchungen metallische Implantate auf.

Nicht alle personenbezogenen Daten konnten für jede Patientin/jeden Patienten ermittelt werden. Besonders die retrospektive Erhebung von Größe und Gewicht der Patientinnen/Patienten zum Zeitpunkt der Untersuchungen gestaltete sich schwierig. Da die Messung von Größe und Gewicht nicht immer am Tag der MRT-Untersuchung durchgeführt wurde, wurde in diesen Fällen die zeitlich am nächsten gelegenen Werte in die Liste aufgenommen. Der Abstand zur MRT-Untersuchung durfte sechs Monate allerdings nicht überschreiten, das Datum der Messung wurde ebenso angegeben. Bei allen Daten zu Größe und Gewicht wurde die Quelle angegeben, manchmal waren auch mehrere Messungen aus unterschiedlichen Quellen verfügbar. Dann wurden beide Quellen und, wenn die Werte voneinander abwichen, auch beide Messergebnisse angegeben. Mögliche Quellen umfassten in den meisten Fällen die Dokumente der Kinderchirurgie wie Aufnahmebögen und Fieberkurven oder Anästhesiedokumentation. In selteneren Fällen waren Dokumente der verschiedenen Ambulanzen der Universitätsklinik für Kinder- und Jugendheilkunde verfügbar. Seit 2014 wird durch die Kinderradiologie ein Formular für jede MRT-Untersuchung angelegt, das auch Größe und Gewicht der Patientin/des Patienten enthält. Für 70 Untersuchungen war ein solches Formular verfügbar.

In den meisten Fällen ist allerdings davon auszugehen, dass diese Werte anamnestisch erhoben wurden. Bei Patientinnen/Patienten, die zum Zeitpunkt der Messung stationär waren, wurden Größe und Gewicht vermutlich durch die Pflege genau erhoben. Dennoch wurde für 98 Datensätze keine Messung innerhalb von sechs Monaten zur MRT-Untersuchung dokumentiert.

3.2 Registrierung und Auswahl der Daten

Bei 86 Patientinnen/Patienten konnte die Registrierung einer MRT-Aufnahme erfolgreich durchgeführt werden. Da nicht garantiert werden konnte, dass der Vorgang bei allen Daten erfolgreich sein würde, wurde in dieser Pilotstudie nur eine MRT-Aufnahme pro Patientin/Patient verwendet. Von den MRT-Aufnahmen, die erfolgreich registriert werden konnten, wurden jene mit Artefakten oder erhöhtem Noise ausgeschlossen. Lagen dann von einer Patientin/einem Patienten immer noch mehrere MRT-Aufnahmen vor, wurde die jeweils jüngste Aufnahme für die SA-Bestimmung verwendet. Das mittlere Alter der 86 Patientinnen/Patienten zum Zeitpunkt der Untersuchung betrug 14,24 Jahre. Drei Patientinnen/Patienten unter zwölf Jahren wurden von der Studie ausgeschlossen, da die Altersgruppe unterrepräsentiert war und einen zu großen Altersunterschied aufwies. Das Alter der übrigen 83 Patientinnen/Patienten lag zwischen 12 und 18 Jahren.

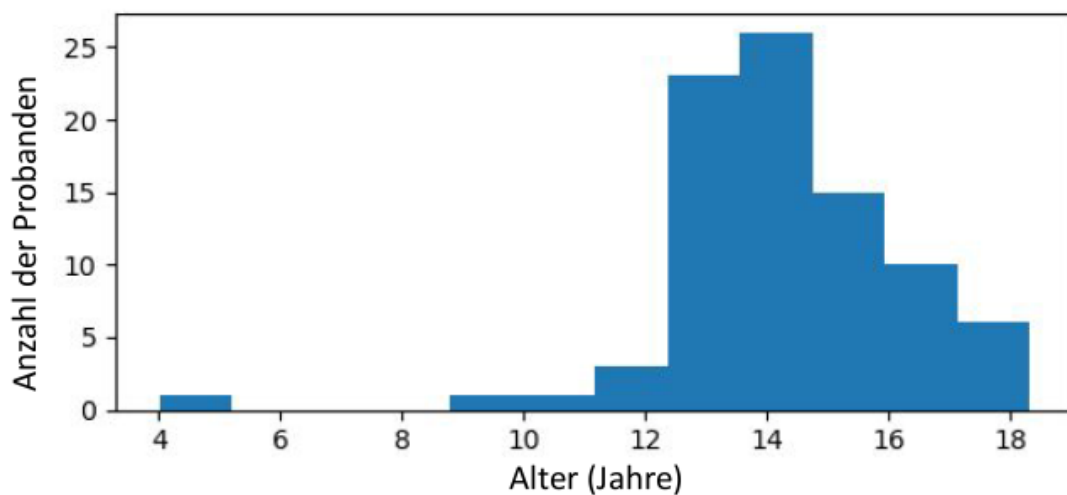


Abbildung 12: Altersverteilung der Studienpopulation nach Registrierung (von LBI CFI zur Verfügung gestellt).

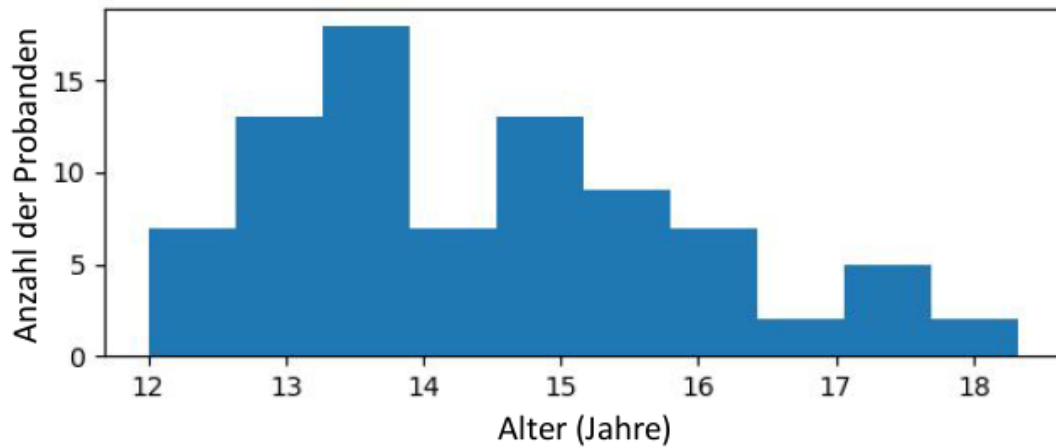


Abbildung 13: Altersverteilung der Studienpopulation nach Registrierung und Ausschluss dreier Patientinnen/Patienten (von LBI CFI zur Verfügung gestellt).

Die 83 Fälle wurden in ein Trainingsset (56 Datensätze) und ein Validierungsset (27 Datensätze) geteilt. Es wurde auf eine möglichst gleiche Verteilung von Alter und Geschlecht geachtet.

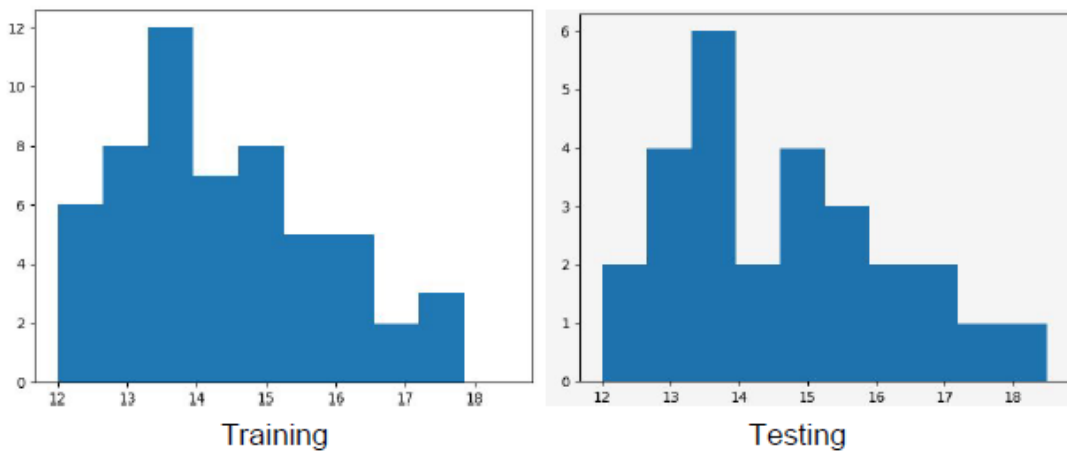


Abbildung 14: Altersverteilung von Trainings- und Validierungsset (von LBI CFI zur Verfügung gestellt).

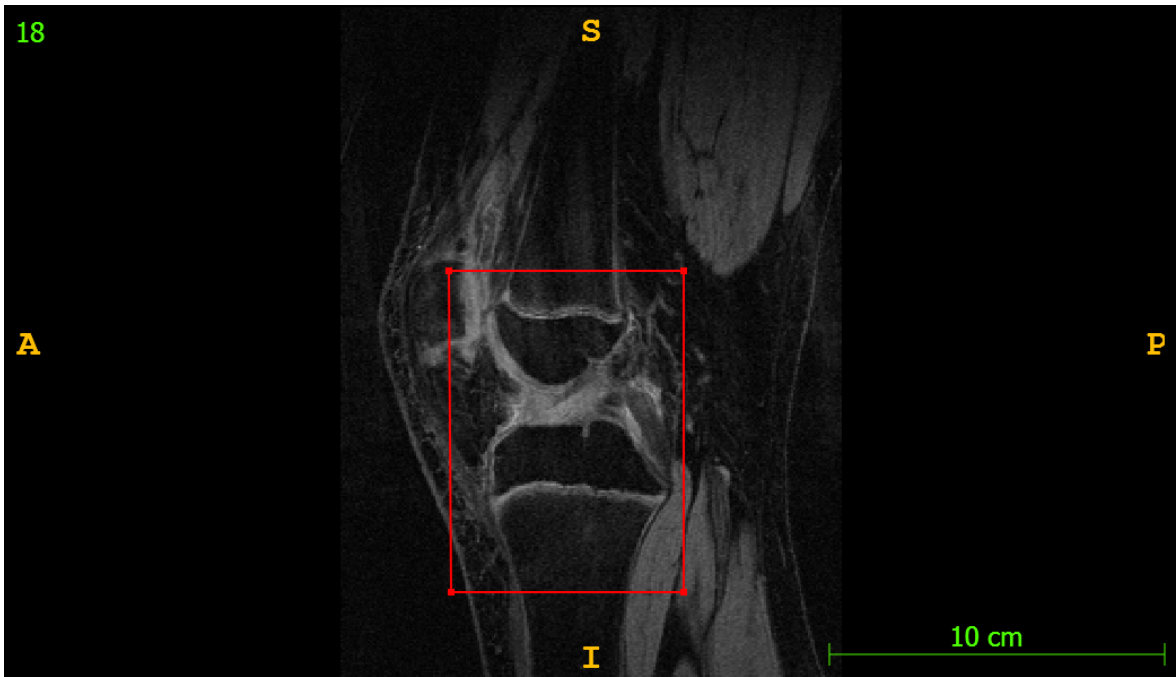


Abb. 15: Beispiel einer MRT-Aufnahme nach erfolgter Registrierung mit rot eingezeichneter ROI bzw. Volume of Interest (von LBI CFI zur Verfügung gestellt).

3.3 Ergebnisse der Altersschätzung

In Experiment 1 wurden Daten von männlichen und weiblichen Patienten gemischt verwendet, da nur 83 Datensätze verfügbar waren. Im elften Durchlauf wurde mit einer absoluten Altersdifferenz von 1,09 Jahren das beste Ergebnis erzielt. In der Validierung kam es zu keiner Verbesserung der Resultate.

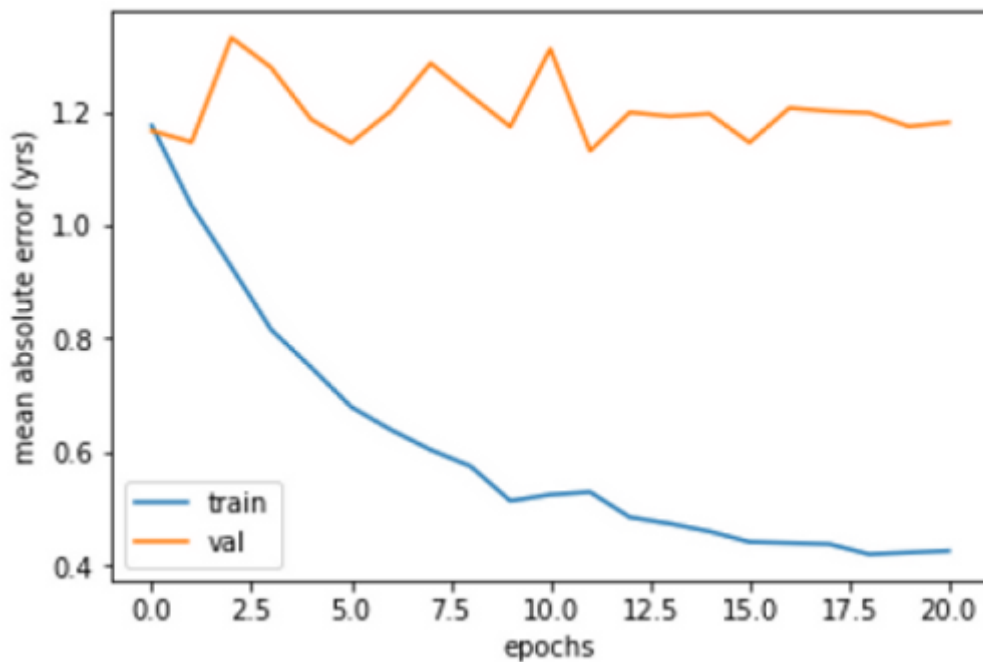


Abbildung 16: Veränderung des mittleren absoluten Fehlers (MAE) in Experiment 1 in Training (blau) und Validierung (orange). Es zeigt sich keine signifikante Verbesserung in der Validierung (von LBI CFI zur Verfügung gestellt).

In den Experimenten 2 und 3 wurden die Datensätze männlicher und weiblicher Patienten getrennt ausgewertet. Die Altersverteilung der wenigen Datensätze wurde manuell angepasst. Der mittlere absolute Fehler („mean absolute error“, MAE) verringerte sich im Training zwar schneller als in Experiment 1, in der Validierung trat jedoch keine Verbesserung ein.

Bei Experiment 4 handelte es sich um ein kontrolliertes Overfitting-Experiment. Dabei wurde die Existenz von für die SA-Bestimmung notwendiger Information in der MRT-Aufnahme überprüft. Die Datensätze männlicher Patienten wurden ohne Augmentation verwendet. Das Resultat ist in Abbildung 17 ersichtlich. Der MAE näherte sich dem Nullwert an, mit einer minimalen Altersdifferenz von 0,13 Jahren im Training.

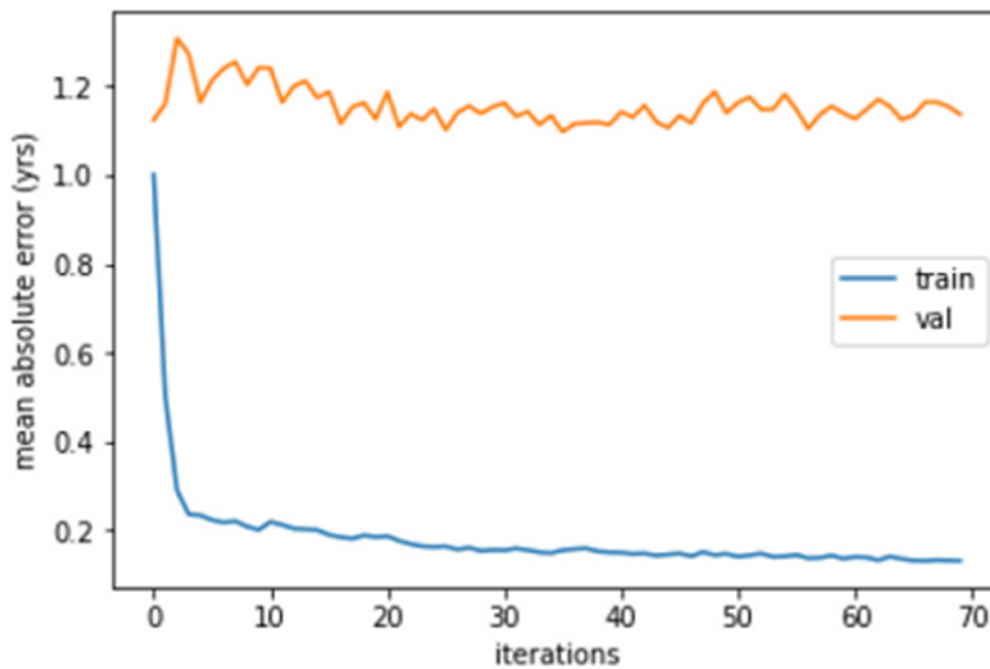


Abbildung 17: Veränderung des MAE in Experiment 4 (kontrolliertes Overfitting-Experiment) in Training (blau) und Validierung (orange)(von LBI CFI zur Verfügung gestellt).

In Experiment 5 wurden ein modifiziertes Residual Network (ResNet42) und alle verfügbaren Datensätze männlicher und weiblicher Patienten verwendet. Mit dieser Architektur ergab sich ein stabilerer Abfall des MAE als zuvor. Sowohl im Training als auch in der Validierung wurden die Ergebnisse präziser. Der niedrigste MAE betrug im Training 0,49 Jahre, in der Validierung 1,15 Jahre.

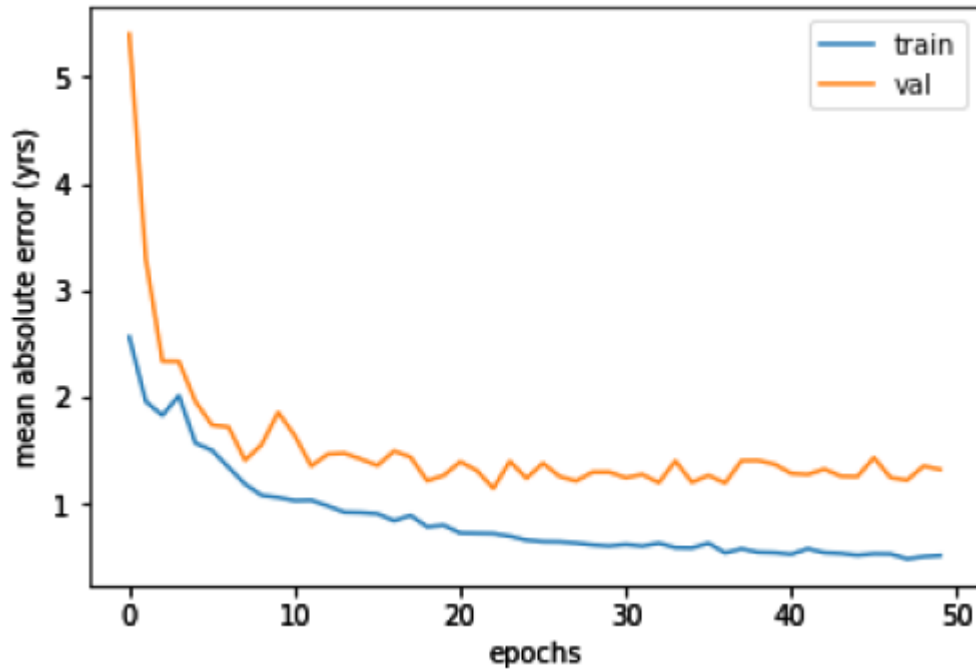


Abbildung 18: Veränderung des MAE in Experiment 5 in Training (blau) und Validierung (orange). Es zeigte sich ein stabiler Abfall des MAE (von LBI CFI zur Verfügung gestellt).

In den Experimenten 6 und 9 wurden dieselben Datensätze wie schon in Experiment 5 verwendet. Initial zeigte sich ein steiler Abfall des MAE. Der geringste MAE lag im Training bei 0,55 Jahren in Experiment 6 und bei 0,42 Jahren in Experiment 9. In der Validierung betrug er in beiden Experimenten 1,04 Jahre. Das ist der niedrigste MAE aller bisher durchgeführten Experimente. In beiden Experimenten kam es jedoch zu einem oder mehreren Ausreißern.

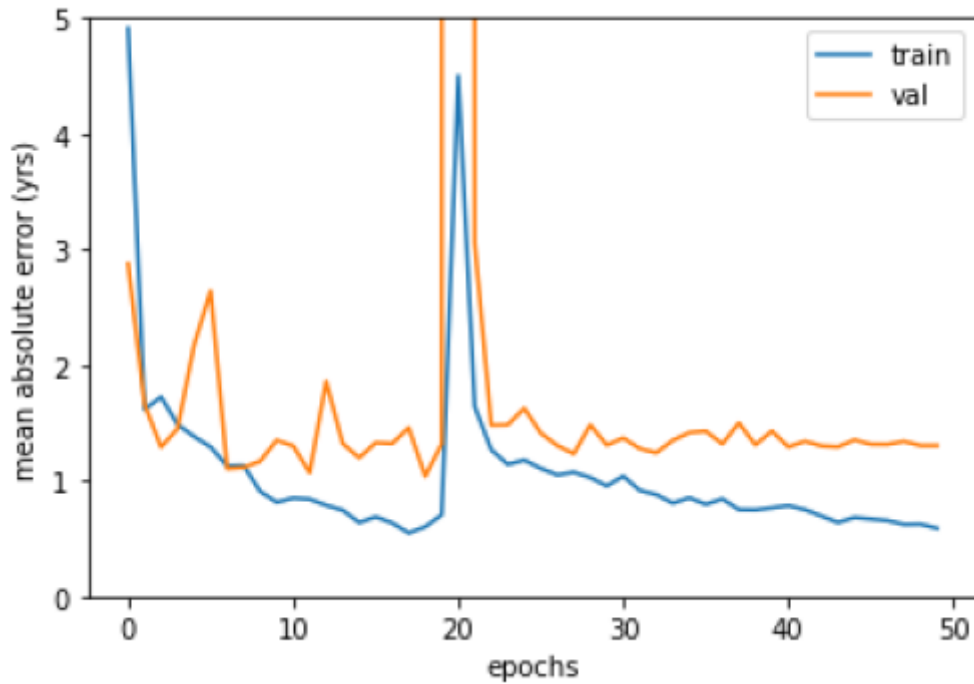


Abbildung 19: Veränderung des MAE in Experiment 6 in Training (blau) und Validierung (orange). Es zeigte sich ein Ausreißer um den 20. Durchlauf, der niedrigste MAE wurde kurz davor erreicht (von LBI CFI zur Verfügung gestellt).

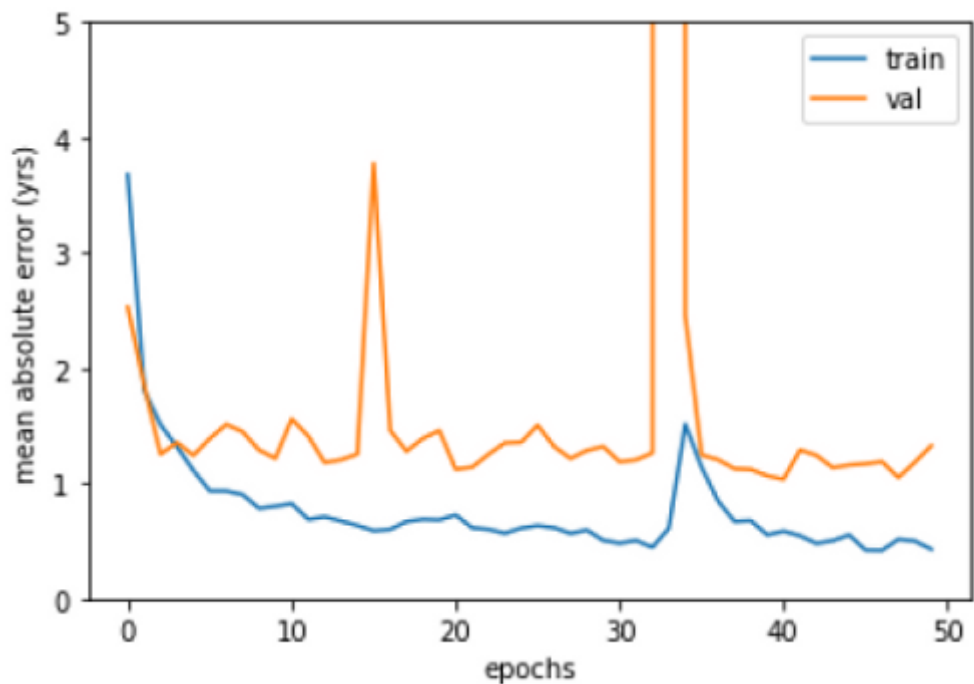


Abbildung 20: Veränderung des MAE in Experiment 9 in Training (blau) und Validierung (orange). Es zeigten sich Ausreißer um 15. sowie zwischen 30. und 40. Durchlauf, niedrigster MAE erst um den 40. Durchlauf erreicht (von LBI CFI zur Verfügung gestellt).

Die Experimente 2, 3, 7, 8, 10 und 11 wiesen keine herausragenden Ergebnisse auf, weshalb sie an dieser Stelle nicht alle grafisch dargestellt werden. In Experiment 8 wurden alle verfügbaren Datensätze von Mädchen und Knaben mit Augmentation verwendet. Wie auch in Experiment 5 handelte es sich um ein modifiziertes Residual Network (ResNet42). Eine Abnahme des MAE blieb aus. Abbildung 21 zeigt einen Vergleich zwischen den Experimenten 5 und 8.

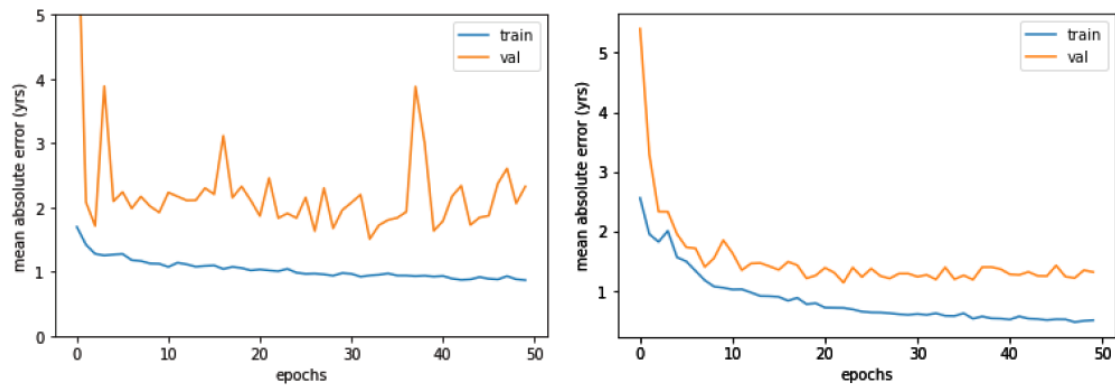


Abbildung 21: Veränderung des MAE in den Experimenten 8 (links) und 5 (rechts) in Training (blau) und Validierung (orange)(von LBI CFI zur Verfügung gestellt).

4 Diskussion

4.1 Limitationen bestehender Methoden und neue Ansätze

4.1.1 Methoden zur Skelettalterbestimmung

Obwohl zahlreiche Methoden zur SA-Bestimmung entwickelt wurden, haben sich nur die wenigsten im klinischen Alltag wirklich durchgesetzt (17). Diese etablierten Methoden wurden zwar im Laufe der Zeit an unterschiedlichsten Studienpopulationen erprobt, weisen jedoch immer noch gewisse Limitationen auf (2, 15, 24-32). Bei der Handskelettalterbestimmung nach Greulich und Pyle umfassen diese nicht nur die bereits beschriebenen Probleme der Subjektivität, Strahlenhygiene, Unsicherheiten bei jüngeren Kindern und Anwendung in anderen Populationen (2, 17). Die Studienpopulation wurde ab dem Alter von fünf Jahren nur einmal jährlich untersucht (2). Ein Jahr ist im präpubertären Wachstumsschub ein sehr großes Zeitfenster. Das bedeutet auch, dass die halbjährlichen Standards, also z.B. 10,5 oder 11,5 Jahre, vom eigentlichen CA der Kinder abweichen und nicht wirklich zu diesem Zeitpunkt aufgenommen wurden. Das Kind, dessen Röntgenbild für den Standard von 12,5 Jahren ausgewählt wurde, hatte zum Zeitpunkt der Untersuchung also nicht das CA von 12,5 Jahren, sondern 12 oder 13. Die Auswahl der Röntgenbilder für den Atlas erfolgte durch die Autoren. Dabei wurden für jedes Alter 100 Bilder vom reifsten bis zum unreifsten Befund gereiht, ein repräsentatives Bild aus der Mitte ausgewählt und in den Atlas aufgenommen. Allerdings entschieden sich die Autoren auch oft für ein Bild, das nicht dem exakten „Mittelwert“ entsprach, um möglichst viele aufeinanderfolgende Bilder desselben Kindes zu verwenden, sofern der SA-Befund dies zuließ. Die in dem Atlas enthaltenen Standards stammen also nur von einer kleinen Gruppe an Probandinnen/Probanden (2).

Mehrere Studien berichten von einem unterschiedlichen SA in Hand und Knie (51, 52). Ob diese Abweichungen der Verwendung unterschiedlicher Methoden geschuldet sind, oder ob tatsächlich Unterschiede in der Geschwindigkeit der Entwicklung der untersuchten Knochen vorliegen, ist fraglich. Der Knieatlas von Pyle und Hoerr ähnelt dem Atlas nach Greulich und Pyle stark in seiner Basis und teilt daher viele seiner Limitationen (2, 50).

4.1.2 Methoden zur Voraussage der Erwachsenengröße

Viele Methoden zur Bestimmung des verbleibenden Längenwachstums beziehen das SA in die Berechnungen mit ein (12, 18, 76, 77). Bis zu einem gewissen Grad sind sie daher durch die vorangegangene SA-Bestimmung limitiert. Eine besondere Herausforderung stellt die Anwendung an jüngeren Kindern dar, da der Wachstumsabschluss in diesen Fällen noch weit entfernt ist (1). Wie auch bei der SA-Bestimmung basieren viele der heute gebräuchlichen Methoden auf nicht mehr aktuellen Studienpopulationen (12, 13, 84).

4.1.3 Neue Ansätze und vergleichbare Studien

Aus oben genannten Gründen werden immer wieder neue Methoden der SA-Bestimmung und Voraussage der Erwachsenengröße entwickelt. Gänzlich neue Ansätze sind jedoch eher selten und nur wenige ziehen dazu das Kniegelenk in Kombination mit der MRT heran (56-61, 64). Diese wenigen Arbeiten stützen sich methodisch in erster Linie auf eine systematische Vermessung oder Einteilung der morphologischen Parameter, was voraussichtlich wiederum in eine Art der Atlaserstellung mit vor allem subjektiven Fehlerquellen in der Anwendung mündet. Weiters sind manuelle Vermessungen zeitaufwendiger als automatisierte Methoden. Die Tendenz zur Automatisierung der SA-Bestimmung erreichte jedoch bisher hauptsächlich Methoden, die Handröntgenbilder verwenden, andere Lokalisationen wie beispielsweise das Kniegelenk oder alternative Bildmodalitäten sind selten vertreten (63, 70, 72-74, 92). Der Vorteil der Objektivität wird vielfach unterschätzt: Intra- und Interobserverabweichungen vollständig zu eliminieren bedeutet im Extremfall einen Gewinn an Präzision von sechs Monaten oder mehr sowie stabilere Ergebnisse (2, 24, 30, 70). Derartige Methoden erfordern lediglich eine Einschulung in das Programm und sind unabhängig von der Erfahrung der Anwenderin/des Anwenders.

Die in dieser Studie untersuchte Methode vereint die Objektivität und Zeitersparnis der Automatisierung mit den Eigenschaften der MRT. Nachteilig an der MRT ist die Notwendigkeit der Sedierung jüngerer Kinder. Allerdings sind Patientinnen/Patienten, bei denen eine SA-Bestimmung durchgeführt wird, in der Regel schon älter und bei vielen Indikationen wird ohnehin eine MRT-Untersuchung durchgeführt. Die Vorteile der Strahlenhygiene und dreidimensionaler Information stehen daher im Vordergrund. Das

Neural Network ist in der Lage, sich mit wachsender Datensammlung zu verbessern (87). Die Daten, die dazu verwendet werden, stammen von österreichischen Kindern und sind aktuell. Dies ist eine weitere Besonderheit der Studie und von großer Bedeutung, da eine Methode in derselben bzw. einer ähnlichen Population meist die genauesten Ergebnisse erzielt. Unabhängig davon, wie Unterschiede des SA zwischen Hand und Knie zustande kommen, ist die Bestimmung des SA bei nicht-forensischen Indikationen dort am sinnvollsten, wo es einen wesentlichen Einfluss auf das Längenwachstum hat. Auch bei der Voraussage der Erwachsenengröße ergeben sich somit Vorteile gegenüber den bestehenden Methoden. Anders als bei der Multiplier Method, die das CA verwendet, können Abweichungen des SA vom CA berücksichtigt werden (13).

Aufgrund der Einzigartigkeit und des frühen Stadiums dieser Studie ist es problematisch, Ergebnisse mit anderen Studien zu vergleichen. BoneXpert ist die erste kommerzielle Software zur automatisierten SA-Bestimmung (70). Wie bereits beschrieben bestimmt BoneXpert das SA an Handröntgenaufnahmen nicht direkt und lehnt Bilder mit viel Noise ab, überzeugt allerdings durch seine Genauigkeit. Eine Voraussage der Erwachsenengröße kann auf Basis der SA-Schätzung ebenso erfolgen (70, 77). Das Alter wird von 2,5 bis 17 Jahren für Knaben und von 2 bis 15 Jahren für Mädchen bestimmt. Die Features der Software wurden manuell festgelegt, weshalb ihr ein selbständiges „Lernen“ nicht möglich ist (70, 73).

Lee et al. entwickelten ein Deep Learning Neural Network zur SA-Bestimmung mittels Handröntgenaufnahmen (73). Dazu wurde eine große Anzahl an Bildern von Patientinnen/Patienten im Alter von fünf Jahren bis zur skelettalen Reife verwendet. CA sowie das SA nach Greulich und Pyle zum Zeitpunkt der Röntgenaufnahmen waren in allen Fällen bekannt, da die Indikation der Handröntgenuntersuchung die SA-Bestimmung war und die Bilder dementsprechend befundet worden waren. In den meisten Fällen lagen Fragestellungen wie verzögerte Pubertät, Pubertas præcox oder Kleinwuchs vor. Anders als in dieser Pilotstudie wurde bei Lee et al. durch den Einsatz eines neuartigen Convolutional Neural Networks (CNN) auch das Preprocessing automatisch durchgeführt, die Methode ist vollautomatisch. Ein weiterer Unterschied zu unserer Pilotstudie ist die Verwendung des SA nach Greulich und Pyle anstelle des CA (73). Alle Ergebnisse des Neural Network basieren somit aber auf SA-Schätzungen von Radiologinnen/Radiologen,

die den Atlas nach Greulich und Pyle verwendeten. Es ist daher fraglich, ob die Software indirekt auch nur eine vollautomatische Atlasvergleichsmethode darstellt oder ob sie über die Limitationen derselben hinauswachsen kann. Fest steht, dass die subjektive Komponente der vorangegangenen SA-Schätzungen auf diese Weise immer noch in der Methode verankert ist. In dieser Pilotstudie hingegen wurde dem Neural Network das CA angegeben, weshalb die Methode völlig unabhängig von dem SA nach Greulich und Pyle ist.

4.2 Limitationen und Stärken der Studie

4.2.1 Ergebnisse der Pilotstudie

Nach Anwendung der Ausschlusskriterien und Registrierung der MRT-Aufnahmen lag eine Datensammlung mit 83 Datensätzen vor. Diese wurde zur Durchführung der SA-Bestimmung herangezogen, was in der Validierung einen niedrigstmöglichen MAE von 1,04 Jahren ergab. Obwohl die Zahl der akquirierten Datensätze zunächst höher als erwartet war, fielen die Altersgrenzen mit 12 bis 18 Jahren relativ eng aus. Gründe dafür waren die Unterrepräsentation jüngerer Altersgruppen, die zum Ausschluss jener Datensätze führte, und die Entscheidung, nur eine MRT-Aufnahme pro Patientin/Patient zu verwenden. Es wäre denkbar, dass die Altersschätzung von 12 bis 18 Jahren statistisch gesehen einfacher ist als jene von 5 bis 18 Jahren. Der größtmögliche Fehler der Software bei Altersgrenzen von 12 bis 18 Jahren beträgt sechs Jahre. Würden die Altersgrenzen jedoch von 5 bis 18 Jahre reichen, wäre eine maximale Abweichung von 13 Jahren möglich. Ein Fehler von sechs Jahren trägt weniger zur Erhöhung des MAE bei als eine Abweichung von 13 Jahren. Relevant ist dies allerdings nur, wenn die Software zufällige Ergebnisse liefert. Dies war in den erfolgreichen Experimenten nicht der Fall und es wurden auch keine derart großen Abweichungen beobachtet. Für die aktuelle Pilotstudie, in deren Mittelpunkt die erste Einschätzung des Lernerfolges des Neural Network an MRT-Aufnahmen des wachsenden Kniegelenks steht, sind diese Altersgrenzen daher ausreichend.

Eine mit dieser Arbeit vergleichbare Studie zur SA-Bestimmung wäre zur Bewertung der Resultate wünschenswert, jedoch bestehen große Unterschiede zwischen den gewählten Methoden der bisher durchgeführten Studien und dieser Pilotstudie. Die an dieser Stelle angestellten Vergleiche der Ergebnisse sind daher nur bedingt aussagekräftig. Aufgrund der weit verbreiteten klinischen Anwendung entspricht die Handskeletalterbestimmung nach Greulich und Pyle noch am ehesten einem Goldstandard der Altersschätzung (2). Daher wird diese zum Vergleich der Altersschätzung an dieser Stelle herangezogen. Glücklicherweise liegt eine Studie vor, die das SA österreichischer Kinder mittels der Methode nach Greulich und Pyle bestimmte (24). Die Abweichungen des geschätzten SA von dem CA beliefen sich auf $-1,5 \pm 7,6$ Monate bei erfahrenen und $2,7 \pm 10,3$ Monate bei weniger erfahrenen Anwenderinnen/Anwendern der Methode (24).

Zum Vergleich unserer Ergebnisse mit den bisherigen automatisierten Methoden zur SA-Bestimmung wurden BoneXpert, eine präzise und ausgereifte Methode, und die aktuelle, vollautomatische Software von Lee et al. ausgewählt (70, 73). BoneXpert wurde an zahlreichen Studienpopulationen validiert (70). Die Genauigkeit wurde als RMSD von der manuellen Handskelettalterbestimmung nach Greulich und Pyle definiert. Geht man davon aus, das SA nach Greulich und Pyle entspräche bei gesunden Kindern dem CA, wäre mit einer RMSD von 0,59 Jahren (7,08 Monate) bis 0,72 Jahre (8,64 Monate) zu rechnen (70).

Die von Lee et al. vorgestellte Software wurde anhand von vier Modellen getestet (73). Die RMSD reichte von 0,82 Jahren bis 1,51 Jahre. Das Modell mit den genauesten Ergebnissen erzielte eine RMSD von 0,93 Jahren (11,16 Monate) bei Mädchen und 0,82 Jahren (9,84 Monate) bei Knaben. Die „mean average precision“ (mAP, mittlere durchschnittliche Präzision) dieses Modells betrug 53,3% bei Mädchen und 55,8% bei Knaben (73).

In unserer Pilotstudie konnte in den Experimenten 6 und 9 mit 1,04 Jahren der geringste MAE erzielt werden. Da es jedoch in beiden Experimenten zu einem oder mehreren großen Ausreißern kam, ist das Ergebnis aus Experiment 5 als das beste und belastbarste Ergebnis zu werten. Dort zeigte sich ein besonders initial steiler Abfall des MAE, der bis auf einen geringen Anstieg um den 9. Durchlauf auch stabil verlief. Der niedrigste MAE betrug 1,15 Jahre. Dies ist sehr nah am Idealfall einer anfangs steil abfallenden Kurve ohne Ausreißer, die im Verlauf flacher wird und sich der Trainingskurve annähert bis schließlich kaum noch eine Verbesserung der Ergebnisse eintritt. Der Kurvenverlauf von Experiment 5 spricht daher für einen Lernfortschritt der Software. Ein gutes Beispiel für das Aussehen einer Kurve ohne Abnahme des MAE als Zeichen eines fehlenden oder geringen Lernfortschritts stellt Experiment 8 dar. Dabei wurden alle verfügbaren Datensätze von Mädchen und Knaben mit Augmentation verwendet. Wie auch in Experiment 5 handelte es sich um ein modifiziertes Residual Network (ResNet42). Es zeigte sich nur ein geringer Abfall des MAE im Training, die Ergebnisse der Validierung wirken zufällig und eine Verbesserung bleibt aus.

Bei Experiment 4 handelte es sich um ein kontrolliertes Overfitting-Experiment, um die Existenz von für die SA-Bestimmung notwendiger Information in der MRT-Aufnahme zu

überprüfen. Das erfolgreiche Overfitting am Trainingsdatensatz ist in Abbildung 17 ersichtlich. Der MAE näherte sich dem Nullwert an, mit einer minimalen Altersdifferenz von 0,13 Jahren im Training. Die Generalisierung war mit dieser Architektur allerdings nicht möglich.

Der MAE von 1,04 Jahren bzw. 1,15 Jahren schneidet im direkten Vergleich zu den genannten Studien zwar schlechter ab, allerdings sollte dabei das frühe Stadium in der Entwicklung der Software berücksichtigt werden (24, 70, 73). Auch die gemischte Verwendung der Datensätze männlicher und weiblicher Patienten könnte den Wert nicht unwesentlich beeinflusst haben. Bezieht man diese Faktoren in die Beurteilung mit ein, scheint ein MAE von 1,15 Jahren ein vielversprechender Anfang zu sein. Es ist sehr wahrscheinlich, dass das Neural Network mit ähnlich großen Datenmengen, wie sie beispielsweise in der Studie von Lee et. al verwendet wurden, präzisere SA-Schätzungen liefern kann (73). Diese Vermutung wird durch den in Experiment 5 ersichtlichen Lerneffekt der Software untermauert.

BoneXpert und die von Lee et al. vorgestellte Software weisen eine gewisse Abhängigkeit von der Methode nach Greulich und Pyle auf (70, 73). Für einen besseren Vergleich mit dieser Pilotstudie wäre die Differenz von SA-Schätzung und CA in einer Population gesunder Kinder notwendig. Verfälschungen der Ergebnisse durch die vorangegangenen manuellen SA-Schätzungen könnten auf diese Weise vermieden werden und ein Vergleich mit unterschiedlichen Methoden, die nicht auf der Handskelettalterbestimmung nach Greulich und Pyle basieren, wäre möglich. Um zukünftige Studien unabhängig von der Methode miteinander vergleichen zu können, sollte daher bei gesunden Kindern, bei denen kein Hinweis auf akzeleriertes oder dezeleriertes SA vorliegt, anstelle des SA nach Greulich und Pyle das CA das Maß für die Genauigkeit bzw. die RMSD sein.

Viel wichtiger als den Vergleich absoluter Werte sehen wir für diese Pilotstudie jedoch den Beweis der Lernfähigkeit der Software und der Wissensgewinn für zukünftige Studien. Die vorliegenden Ergebnisse sind im Vergleich zur aktuellen Literatur die ersten dieser Art, nachdem im Vorfeld auch nicht mit Sicherheit davon ausgegangen werden konnte, ob eine SA-Schätzung überhaupt möglich wäre und wie präzise die Ergebnisse ausfallen würden. Die Hypothese, dass ein objektives Deep Learning Neural Network auf Basis von Algorithmen des maschinellen Lernens und der medizinischen Bildanalyse eine

Schätzung des SA anhand von MRT-Daten des Knies liefern kann, konnte durch diese Studie bestätigt werden. Die Lernfähigkeit konnte durch die Abnahme des MAE in mehreren Experimenten bewiesen werden. In Bereichen wie Studiendesign, Möglichkeiten der Datenakquirierung und technischen Details konnten neue Erkenntnisse gewonnen werden. In dieser Hinsicht sind die Ergebnisse unserer Studie ermutigend und erlauben die Planung weiterführender Studien.

4.2.2 Methodik und Studiendesign

Wie bereits erläutert wurde, kamen bei der vorliegenden Pilotstudie neuartige Methoden zum Einsatz, was mit Herausforderungen verbunden war. Von Seiten des LBI CFI bestanden Vorerfahrungen mit Neural Networks, allerdings bezogen sich diese bisher auf die Anwendung an Wachstumsfugen anderer Knochen (62, 63).

4.2.2.1 Datenakquirierung und Studienpopulation

Für die durchgeführte Bestimmung des SA wurden die Daten zu Größe und Gewicht aus der Datensammlung nicht verwendet. In 98 Fällen konnten in der Recherche dazu keine Daten gefunden werden, der Rest setzt sich aus größtenteils anamnestisch erhobenen Werten ohne Messung der Parameter zum Zeitpunkt der Erhebung zusammen. Die aus Fieberkurven entnommenen Werte sind vermutlich am genauesten, jedoch kann dies nicht überprüft werden. Prospektive Studien mit Messungen von Größe und Gewicht zum Untersuchungszeitpunkt weisen den Nachteil dieser Ungenauigkeiten nicht auf, was für weiterführende Studien hinsichtlich der Voraussage der Erwachsenengröße eine interessante Erweiterung im Informationsgewinn bedeuten könnte. Auch die ethnische Herkunft konnte retrospektiv nicht ermittelt werden. Wie schon durch Schmeling et al. und Sutow et al. gezeigt wurde, besteht allerdings Grund zur Annahme, dass der Einfluss der ethnischen Herkunft auf die Skelettreife in dieser Fragestellung vernachlässigbar ist (7, 93).

Ein Unterschied zu bestehenden Arbeiten findet sich in der Wahl der Ausschlusskriterien der Datensätze. Da es sich bei der Studienpopulation um Patientinnen/Patienten handelt, die aus medizinischen Indikationen einer MRT-Untersuchung zugeführt wurden, liegen in den meisten Fällen Pathologien im Bereich des Kniegelenks vor. Diese waren meist traumatisch bedingt, besonders häufig fanden sich Patellaluxation, Ruptur des vorderen

Kreuzbandes oder Meniskusverletzungen. Viele Patientinnen/Patienten wurden chirurgisch behandelt und wiesen in Verlaufsuntersuchungen metallische Implantate auf. In wenigen Fällen lag beispielsweise die Diagnose einer juvenilen idiopathischen Arthritis vor, was in anderen Studien einen Grund für den Ausschluss der Patientin/des Patienten dargestellt hätte (57). Andere Studien wandten somit strengere Ausschlusskriterien an oder verwendeten die Daten freiwilliger, gesunder Probandinnen/Probanden (57, 74, 94, 95). Allerdings stehen die pathologischen Veränderungen im untersuchten Altersabschnitt nicht in direktem Zusammenhang mit dem Alter der Patientinnen/Patienten. Die Ergebnisse der SA-Schätzung lassen vermuten, dass es dem Neural Network möglich war, dies zu erkennen und stattdessen für das Alter relevante Information zu verwenden. Für unsere Pilotstudie wurden die Ausschlusskriterien daher in dieser Form nahe am klinischen Alltag gewählt, und stellen bewusst eine zusätzliche Herausforderung für die Software dar. Es ist denkbar, dass engere Ausschlusskriterien zu „saubereren“ Datensätzen und somit zu einem geringeren MAE führen könnten.

Die Wahl des Zeitabstandes von sechs Monaten zwischen den MRT-Untersuchungen als Einschlusskriterium erfolgte aufgrund der Überlegung, dass in diesem Zeitfenster kaum relevante Veränderung an den Wachstumsfugen eintreten würde und auf diese Weise vorrangig wachstumsbedingte Veränderungen im Datensatz in der Auswertung durch das Neural Network Niederschlag finden würden. Allerdings wurde nicht berücksichtigt, dass auch minimale Veränderungen oder das Fehlen von Veränderung Informationen liefert und womöglich durch die Software hätte erkannt werden können. Jeder zusätzliche MRT-Datensatz, der erfolgreich registriert werden kann, trägt zur Verbesserung des Neural Network bei. Bei der Augmentation wird dieselbe Aufnahme mehrmals verwendet, obwohl die Wachstumsfugen dieselben sind. Besonders im präpubertären Wachstumsschub könnte der Abstand von sechs Monaten diesbezüglich großzügig gewählt worden sein. Die Wahl des zeitlichen Abstandes zwischen den Zeitpunkten der Erhebung der Datensätze wirft auch die Frage auf, wie präzise eine Altersschätzung überhaupt durchgeführt werden kann und welches angestrebte Maß an Präzision sinnvoll ist. Individuelle Unterschiede zwischen Probandinnen/Probanden, Umwelteinflüsse und viele andere Faktoren erschweren die Beantwortung dieser Frage. Solange dieser Aspekt der Altersschätzung unklar ist, erscheint es für weiterführende zukünftige Studien sinnvoll, einen Zeitabstand eher zu klein als zu groß zu wählen. In prospektiven Studien

verhindern unterschiedliche Gründe die Wahl einer derart großen Anzahl an Messzeitpunkten. Ein retrospektives Studiendesign würde dies aber ermöglichen.

Die Datensammlung und somit auch die für die Altersschätzung verwendeten Daten zeigen eine ungleichmäßige Altersverteilung. Während die Altersgruppe von 13 bis 14 Jahren überrepräsentiert war, lagen nur wenige Datensätze von Patientinnen/Patienten über 15,5 Jahren vor. Die randomisierte Zuordnung in Trainings- und Validierungsdatensatz wurde für die einzelnen Experimente manuell optimiert. Würde das Neural Network jeden Datensatz auf 13 oder 14 Jahre schätzen, ohne für das Alter relevante Strukturen identifiziert zu haben, läge der MAE niedriger als in einer Studienpopulation mit ausgeglichener Altersverteilung, denn das wahre Alter der untersuchten Patientinnen/Patienten liegt mit hoher Wahrscheinlichkeit zwischen 13 und 14 Jahren. Der MAE allein reicht daher nicht aus, um die Qualität der Ergebnisse zu beschreiben. Die einzelnen Resultate der Experimente sprechen jedoch gegen ein „Raten“ der Software, denn die Schätzungen lagen oft jenseits dieser Altersgrenzen und es konnte im Verlauf ein Lernfortschritt beobachtet werden. Für zukünftige Experimente ist eine Studienpopulation mit gleichmäßiger Altersverteilung jedoch vorzuziehen, um diesem Bias vorzubeugen.

Die Beurteilung des Wachstumsfugenschlusses erfolgte subjektiv mit Ausschluss der auf den ersten Datensatz mit geschlossener Fuge nachfolgenden Datensätze. Da im Falle einer subjektiv geschlossenen Wachstumsfuge des distalen Femur kein Längenwachstum zu erwarten ist, wurde das Vorgehen für die vorliegende Fragestellung passend gewählt (56). Allerdings liegt auch eine Studie vor, die Veränderungen der Wachstumsfugen des Kniegelenks bis weit über das 18. Lebensjahr hinaus postuliert (58). Für die Altersbestimmung Jugendlicher und Erwachsener aus forensischer Indikation sind diese Veränderungen relevant. Durch den Ausschluss der beschriebenen MRT-Aufnahmen wurde dem Neural Network die Möglichkeit genommen, jene Veränderungen zu erkennen. Der in dieser Pilotstudie getestete Anwendungsbereich der Software war diesbezüglich auf die Fragestellung der Wachstumsschätzung eingeschränkt.

4.2.2.2 Schätzung des SA

Prinzipiell gilt: Je mehr Daten einem Neural Network zur Verfügung gestellt werden, desto genauer fallen seine Ergebnisse aus (87, 88). Daher ist es bedauerlich, dass von den ursprünglichen 317 MRT-Aufnahmen nur 86 erfolgreich registriert werden konnten. Im Vergleich zu anderen Studien mit ähnlichem Aufbau ist das eine kleine Studienpopulation: Lee et al. beispielsweise verwendeten zur vollautomatischen SA-Bestimmung über 8000 Handröntgenaufnahmen (73). Pro Patientin/Patient wurde ein einziger MRT-Datensatz ausgewertet, um der Entstehung eines Bias durch die Verwendung einer unterschiedlichen Anzahl von MRT-Aufnahmen pro Patientin/Patient vorzubeugen. Für weiterführende Untersuchungen böte sich die Verwendung mehrerer Datensätze pro Patientin/Patient an, sofern die Verteilung gleichmäßig bleibt und eine erfolgreiche Registrierung der MRT-Aufnahmen gewährleistet werden kann. Es könnte auch eine weitere Datensammlung mit den Daten jener Patientinnen/Patienten, die nur eine MRT-Untersuchung erhielten, angelegt und zur SA-Bestimmung herangezogen werden. Zu ergründen, warum die Registrierung zahlreicher Datensätze nicht erfolgreich durchgeführt werden konnte, würde die genaue Überprüfung sämtlicher MRT-Aufnahmen erfordern. Da dies einen hohen zeitlichen Aufwand darstellt, konnte bisher nur eine Stichprobe überprüft werden. Als Gründe für den Fehlschlag der Registrierung werden in diesen Fällen erhöhtes Noise, Pathologien und zu starke Abweichungen des „Field of view“ von der manuell registrierten Aufnahme angenommen. Die Stichprobe reicht allerdings noch nicht aus, um die Frage nach der Ursache zufriedenstellend zu beantworten.

Bei der Definition der ROI besteht voraussichtlich noch die Möglichkeit zur Verbesserung. Die ROI wurde in dieser Studie anhand einer MRT-Aufnahme manuell eingegeben, die übrigen Aufnahmen wurden dann auf diese registriert. Lee et al. hingegen verwendeten ein Detection Convolutional Neural Network (CNN), das die ROI automatisch erkannte (73). Das liegt in erster Linie daran, dass dies für zweidimensionale Röntgenbilder einfacher zu verwirklichen ist als für dreidimensionale Daten.

Die geringe Anzahl an Datensätzen führte nach den ersten Experimenten zu der Entscheidung, männliche und weibliche Patienten zu einer Gruppe zusammenzuschließen. Die Information über das Geschlecht wurde dem Neural Network

nicht zur Verfügung gestellt. In Anbetracht der Tatsache, dass Mädchen in der Regel ein höheres SA als Knaben aufweisen, ist es verwunderlich, dass in den Experimenten 6 und 9, in denen Daten von Mädchen und Knaben gemischt verwendet wurden, ein MAE von 1,04 Jahren erzielt werden konnte (4). Es wäre jedoch möglich, dass das Neural Network durch die Erhöhung der Datenmenge auch die Unterscheidung zwischen Mädchen und Knaben lernen konnte. Allerdings erschwert diese weitere Herausforderung das Training, weshalb die Bereitstellung dieser Information in zukünftigen Studien sinnvoll ist.

Aufgrund der hohen Verfügbarkeit und Auflösung der T2-gewichteten MRT-Aufnahmen wurde diese Sequenz für die Altersschätzung einheitlich verwendet. Andere Sequenzen lagen weniger regelmäßig vor. Krämer et al. hingegen beschrieben die T1-gewichtete Sequenz aufgrund der indirekten Darstellung der Knochen als gut geeignet für die Erkennung früher Stadien der Ossifikation (57, 58). Allerdings besteht bezüglich der Wahl der Sequenz kein Konsens und andere Autorinnen/Autoren bevorzugten die T2-Gewichtung (56, 64). Ekizoglu et al. verglichen in einer retrospektiven Studie die Sequenzen miteinander und wählten die T2-Gewichtung mit der Begründung, in dieser Sequenz wären gewisse anatomische Strukturen, hyperintense Überreste von Knorpelgewebe und Pathologien wie Knochenmarksödeme besser sichtbar (64). Es ist unklar, ob und wie die Wahl der Sequenz die Ergebnisse beeinflusst hat. Auch für die Wahl der Schnittebene existiert keine eindeutige Empfehlung. Wie auch in den Studien von Krämer et al. wurden Aufnahmen in sagittaler Schnittebene verwendet (57, 58). Anders als bei der Erstellung eines zweidimensionalen Atlas hat die Wahl der Schnittebene in dieser Pilotstudie aber keinen Einfluss auf das Ergebnis. Grund dafür ist, dass in der gewählten Sequenz die Auflösung nahezu isotrop ist und alle Prozesse nicht in einzelnen Ebenen, sondern in dreidimensionalen Volumina erfolgen. Bei dem verwendeten MRT-Gerät handelte es sich um ein Modell mit 1,5 Tesla. Ob die durch ein Gerät mit höherer Flussdichte erzielbare verbesserte Bildqualität zu präziseren Ergebnissen führt, kann noch nicht beantwortet werden.

Zur selbständigen Arbeitsweise eines Neural Network muss zur Diskussion gestellt werden, dass die einzelnen Schritte und Bezugspunkte der automatisierten Berechnungen schwer einsehbar oder definierbar sind. Bei der SA-Schätzung liefert das Neural Network das Ergebnis in Form einer Zahl. Da dieses errechnete Alter die einzige angegebene

Information darstellt, ist nicht direkt nachvollziehbar, wie das Ergebnis zustande gekommen ist und was die Software gemessen und berechnet hat. Die Messungen, Lernprozesse und Gewichtungen des Neural Network bleiben verborgen. Allerdings wurden Methoden entwickelt, die einen Einblick in die Arbeitsweise des Neural Network erlauben (96, 97). Eine davon ist die Okklusionsmethode, bei der ein Teil des Bildes bzw. Volumens verdeckt wird und beobachtet wird, wie dies das Ergebnis verändert. Dieser Vorgang wird für das gesamte Bild wiederholt, sodass die relevanten Bereiche ersichtlich werden. Bereiche, denen das Neural Network viel Bedeutung beimisst, können als farbige Markierung über dem Bild dargestellt werden, sodass eine sogenannte Attention Map entsteht. Sind die Wachstumsfugen auf der Attention Map markiert, bedeutet dies, dass das Neural Network an diesen Stellen seine Information bezieht (73, 96). Im Rahmen einer anderen Studie des LBI CFI wurde ein vergleichbares Neural Network auf diese Weise getestet. Die Wachstumsfugen wurden als relevante Bereiche identifiziert, was einem zufriedenstellenden Ergebnis entspricht. Daher ist es naheliegend, dass das Neural Network auch in dieser Pilotstudie die Wachstumsfugen als für die Altersbestimmung ausschlaggebende Bereiche erkennen konnte. Die hier diskutierten, ermutigenden ersten Ergebnisse einer automatisierten Schätzung des SA aus MRT-Aufnahmen des Kniegelenks untermauern diese Vermutung. Für die Absicherung zukünftiger, umfangreicherer Forschungsergebnisse dieser Art ist die Anwendung einer solchen Methode jedenfalls empfehlenswert.

4.3 Erkenntnisse und Schlussfolgerungen

Diese Diplomarbeit bietet eine Einführung in die beiden Themen der SA-Bestimmung und der Voraussage der Erwachsenengröße mit einer Zusammenfassung vorhandener Literatur sowie einer Beschreibung der bekanntesten Methoden. Von den gängigsten Methoden zur SA-Bestimmung ist trotz der in Kapitel 1.2.1.1.2 angeführten Limitationen nach wie vor die Handskelettalterbestimmung nach Greulich und Pyle mit den Korrekturen von Cole et. al für österreichische Kinder am ehesten geeignet (2, 24, 33). Dies ist allerdings auch auf die langjährige Expertise der Anwenderinnen/Anwender mit der Methode zurückzuführen, da die Genauigkeit der Schätzungen mit der Erfahrung zunimmt (2, 24). Eine objektive, an die aktuelle österreichische Population angepasste Methode ohne Strahlenbelastung würde eine bedeutsame Verbesserung in vielen Anwendungsbereichen der SA-Bestimmung darstellen.

Weiters erfolgte der Aufbau einer MRT-Datensammlung, die 317 Datensätze umfasst und auch für zukünftige Studien herangezogen werden kann. Einige der erhobenen Zielgrößen wurden in dieser Pilotstudie noch nicht verwendet, sind jedoch für weiterführende Schritte wie die longitudinale Auswertung der Daten im Sinne einer Voraussage der Erwachsenengröße von Bedeutung.

Bei der Kalibrierung und Validierung des Deep Learning Neural Networks konnten besonders in Bezug auf die Architektur der Software einige wertvolle Erkenntnisse gewonnen werden. Unterschiedliche Architekturen wurden getestet. Zusammenfassend konnte beobachtet werden, dass Residual Neural Networks in der Validierung stabilere Ergebnisse aufwiesen als Plain Network-Architekturen. Allerdings erreichten sowohl Plain Networks als auch Residual Networks denselben minimalen MAE von 1,04 Jahren in der Validierung. Aufgrund der geringen Datenmenge konnte zum Einfluss der Augmentation auf die Ergebnisse keine verlässliche Aussage getroffen werden. Um die Genauigkeit und Aussagekraft der Ergebnisse weiter zu verbessern, ist die Akquirierung größerer Datenmengen unerlässlich.

Dass die zur automatisierten SA-Bestimmung mittels Neural Network notwendige Information in den MRT-Aufnahmen des Knies enthalten ist, konnte in den durchgeführten Experimenten dieser Studie eindeutig nachgewiesen werden. Weiters

zeigte sich ein Lernfortschritt der Software. Die Hypothese, dass die SA-Schätzung aus standardisierten MRT-Aufnahmen des wachsenden Kniegelenks mittels darauf trainierter selbstlernender Software möglich ist, konnte somit bestätigt werden. In Hinblick auf die Präzision steht die Methode jedoch erst am Anfang. Die insgesamt ermutigenden Resultate legen nahe, dass eine deutliche Verbesserung und ein Ausbau der Methode zukünftig durchaus möglich sind.

4.4 Ausblick

Die aus dieser Studie gewonnenen Erkenntnisse erlauben eine Einschätzung der Sinnhaftigkeit nachfolgender Studiendesigns. Das Studienkonzept sah vor, bei feststellbarem Lernerfolg der Software Folgestudien durchzuführen. Ein sinnvoller nächster Schritt wäre die Akquirierung größerer Datenmengen. Eine größere Datensammlung würde voraussichtlich zu stabileren und präziseren Ergebnissen führen. Die in dieser Studie bewährten Architekturen des Neural Network könnten an diesen neuen Daten weiter erprobt und verbessert werden, allerdings sollten auch die diskutierten möglichen Anpassungen in der Methodik in Erwägung gezogen werden. Besonders die Überprüfung und gegebenenfalls Verbesserung der Registrierung bieten sich als einen weiteren Schritt an, da in diesem Bereich noch Potential zu vermuten ist. Sobald ausreichend Daten von Kindern und Jugendlichen aller Altersgruppen vorhanden sind, können Altersschätzungen in einem größeren Altersumfang getestet werden. Ein weiterer Schritt wäre die Nutzung der gesamten bestehenden Datensammlung mit longitudinalen Daten zur Voraussage der Erwachsenengröße. Da diese Fragestellung in der Pilotstudie nicht behandelt wurde, ist derzeit noch unklar, ob das Neural Network die zusätzliche Aufgabe der Prognose bewerkstelligen kann.

Abhängig von weiteren Studienerfolgen und Entwicklungen kann in Zukunft auch eine Studie mit prospektivem Design in Erwägung gezogen werden. Eine kontrollierte und standardisierte Erhebung von Körpergröße, SA und anderen Parametern erscheint vorteilhafter als die rein retrospektive Sammlung dieser Daten.

Es ist denkbar, dass auf diesem Wege eine vollautomatisierte, präzise und strahlungsfreie SA-Bestimmung mit gleichzeitiger Voraussage des verbleibenden Wachstums auf Knopfdruck in Zukunft nicht nur Spekulation bleibt, sondern Einzug in den klinischen Alltag findet.

5 Literaturverzeichnis

1. Exner GU. Normalwerte in Wachstum und Entwicklung: die Basis für Diagnostik und Therapie ; 29 Tabellen: Thieme; 2003.
2. Pyle SI, Greulich WW. Radiographic Atlas of Skeletal Development of the Hand and Wrist. Stanford: Stanford University Press; 1959. 272 p.
3. Simmons K, Greulich WW. Menarcheal age and the height, weight, and skeletal age of girls age 7 to 17 years. *The Journal of Pediatrics*.22(5):518-48.
4. De Sanctis V, Di Maio S, Soliman AT, Raiola G, Elalaily R, Millimaggi G. Hand X-ray in pediatric endocrinology: Skeletal age assessment and beyond. *Indian Journal of Endocrinology and Metabolism*. 2014;18(Suppl 1):S63-S71.
5. Nahhas RW, Sherwood RJ, Chumlea WC, Towne B, Duren DL. Predicting the Timing of Maturational Spurts in Skeletal Age. *American journal of physical anthropology*. 2013;150(1):68-75.
6. Greulich WW. A comparison of the physical growth and development of American-born and native Japanese children. *Am J Phys Anthropol*. 1957;15(4):489-515.
7. Sutow WW. Skeletal Maturation in Healthy Japanese Children, 6 to 19 Years of Age. Comparison with Skeletal Maturation in American Children. *Hiroshima Journal of Medical Sciences*. 1953;2:181-91.
8. Sutow WW, Ohwada, K. Skeletal Standards of Healthy Japanese Children from Age 6 to 19 Years. *Clinical Pediatrics (Japanese)*. 1953;6(No. 11).
9. Proos LA. Anthropometry in adolescence--secular trends, adoption, ethnic and environmental differences. *Hormone research*. 1993;39 Suppl 3:18-24.
10. Martin DD, Wit JM, Hochberg Z, Savendahl L, van Rijn RR, Fricke O, et al. The use of bone age in clinical practice - part 1. *Horm Res Paediatr*. 2011;76(1):1-9.
11. Martin DD, Wit JM, Hochberg Z, van Rijn RR, Fricke O, Werther G, et al. The use of bone age in clinical practice - part 2. *Horm Res Paediatr*. 2011;76(1):10-6.
12. Anderson M, Green WT, Messner MB. The classic. Growth and predictions of growth in the lower extremities by Margaret Anderson, M.S., William T. Green, M.D. and Marie Blail Messner, A.B. from the *Journal of Bone and Joint Surgery*, 45A:1, 1963. *Clinical orthopaedics and related research*. 1978(136):7-21.
13. Paley D. Principles of Deformity Correction. 1 ed: Springer-Verlag Berlin Heidelberg; 2002. XXV, 806 p.
14. D. BR, E. HJ, C. SS, Dror P. Temporary hemiepiphyseal arrest using a screw and plate device to treat knee and ankle deformities in children: a preliminary report. *Journal of Children's Orthopaedics*. 2008;2(3):187-97.
15. Zabet D, Rérolle C, Pucheux J, Telmon N, Saint-Martin P. Can the Greulich and Pyle method be used on French contemporary individuals? *International Journal of Legal Medicine*. 2015;129(1):171-7.
16. De Donno A, Roca R, Introna F, Santoro V. A case of an adoptive girl with precocious puberty: The problem of age estimation. *Forensic Science International*.231(1):400.e1-.e4.
17. Breen MA, Tsai A, Stamm A, Kleinman PK. Bone age assessment practices in infants and older children among Society for Pediatric Radiology members. *Pediatr Radiol*. 2016;46(9):1269-74.
18. Tanner JM. Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method): W.B. Saunders; 2001.

19. Chumela WC, Roche AF, Thissen D. The FELS method of assessing the skeletal maturity of the hand-wrist. *American Journal of Human Biology*. 1989;1(2):175-83.
20. FISHMAN LS. Radiographic Evaluation of Skeletal Maturation. *The Angle Orthodontist*. 1982;52(2):88-112.
21. Satoh M. Bone age: assessment methods and clinical applications. *Clinical Pediatric Endocrinology*. 2015;24(4):143-52.
22. King DG, Steventon DM, O'Sullivan MP, Cook AM, Hornsby VPL, Jefferson IG, et al. Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods. *The British Journal of Radiology*. 1994;67(801):848-51.
23. Euling SY, Herman-Giddens ME, Lee PA, Selevan SG, Juul A, Sørensen TIA, et al. Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends: Panel Findings. *Pediatrics*. 2008;121(Supplement 3):S172-S91.
24. Groell R, Lindbichler F, Riepl T, Gherra L, Roposch A, Fötter R. The reliability of bone age determination in central European children using the Greulich and Pyle method. *The British Journal of Radiology*. 1999;72(857):461-4.
25. Loder RT, Estle DT, Morrison K, Eggleston D, Fish DN, Greenfield ML, et al. Applicability of the Greulich and Pyle skeletal age standards to black and white children of today. *American journal of diseases of children (1960)*. 1993;147(12):1329-33.
26. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *American Journal of Roentgenology*. 1996;167(6):1395-8.
27. Paxton ML, Lamont AC, Stillwell AP. The reliability of the Greulich–Pyle method in bone age determination among Australian children. *Journal of Medical Imaging and Radiation Oncology*. 2013;57(1):21-4.
28. Mughal AM, Hassan N, Ahmed A. The applicability of the Greulich & Pyle Atlas for bone age assessment in primary school-going children of Karachi, Pakistan. *Pakistan Journal of Medical Sciences*. 1969;30(2).
29. Mora S, Boechat MI, Pietka E, Huang HK, Gilsanz V. Skeletal Age Determinations in Children of European and African Descent: Applicability of the Greulich and Pyle Standards. *Pediatr Res*. 2001;50(5):624-8.
30. Tisè M, Mazzarini L, Fabrizzi G, Ferrante L, Giorgetti R, Tagliabracci A. Applicability of Greulich and Pyle method for age assessment in forensic practice on an Italian sample. *International Journal of Legal Medicine*. 2011;125(3):411-6.
31. Santoro V, Roca R, De Donno A, Fiandaca C, Pinto G, Tafuri S, et al. Applicability of Greulich and Pyle and Demirijian aging methods to a sample of Italian population. *Forensic Science International*. 2012;221(1):153.e1-.e5.
32. Wenzel A, Droschl H, Melsen B. Skeletal maturity in Austrian children assessed by the GP and the TW-2 methods. *Ann Hum Biol*. 1984;11(2):173-7.
33. Cole AJ, Webb L, Cole TJ. Bone age estimation: a comparison of methods. *Br J Radiol*. 1988;61(728):683-6.
34. Lynnerup N, Belard E, Buch-Olsen K, Sejrsen B, Damgaard-Pedersen K. Intra- and interobserver error of the Greulich–Pyle method as used on a Danish forensic sample. *Forensic Science International*. 2008;179(2):242.e1-.e6.
35. Horter MJ, Friesen S, Wacker S, Vogt B, Leidiger B, Roedl R, et al. Bestimmung des Skelettalters. *Der Orthopäde*. 2012;41(12):966-76.
36. Beunen G, Lefevre J, Oystyn M, Renson R, Simons J, Van Gerven D. Skeletal maturity in Belgian youths assessed by the Tanner-Whitehouse method (TW2). *Ann Hum Biol*. 1990;17(5):355-76.

37. Labarthe DR, Dai S, Day RS, Fulton JE, Grunbaum JA, Shah SM, et al. Project HeartBeat! Concept, development, and design. *American journal of preventive medicine*. 2009;37(1 Suppl):S9-16.
38. Maggio A, Flavel A, Hart R, Franklin D. Skeletal age estimation in a contemporary Western Australian population using the Tanner–Whitehouse method. *Forensic Science International*. 2016;263:e1-e8.
39. Schmidt S, Nitz I, Schulz R, Schmeling A. Applicability of the skeletal age determination method of Tanner and Whitehouse for forensic age diagnostics. *International Journal of Legal Medicine*. 2008;122(4):309.
40. Cox LA. Tanner-Whitehouse method of assessing skeletal maturity: problems and common errors. *Hormone research*. 1996;45 Suppl 2:53-5.
41. Proos LA, Lönnerholm T, Jonsson B, Tuvemo T. Can the TW3 Bone Age Determination Method Provide Additional Criteria for Growth Hormone Treatment in Adopted Girls with Early Puberty? *Hormone Research in Paediatrics*. 2010;73(1):35-40.
42. Nahhas RW, Sherwood RJ, Chumlea WC, Duren DL. An update of the statistical methods underlying the FELS method of skeletal maturity assessment. *Annals of human biology*. 2013;40(6):505-14.
43. Vignolo M, Milani S, Cerbello G, Coroli P, Di Battista E, Aicardi G. FELS, Greulich-Pyle, and Tanner-Whitehouse bone age assessments in a group of Italian children and adolescents. *American journal of human biology : the official journal of the Human Biology Council*. 1992;4(4):493-500.
44. Javadinejad S, Sekhavati H, Ghafari R. A Comparison of the Accuracy of Four Age Estimation Methods Based on Panoramic Radiography of Developing Teeth. *Journal of Dental Research, Dental Clinics, Dental Prospects*. 2015;9(2):72-8.
45. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. *Human biology*. 1973;45(2):211-27.
46. Risser JC. The Classic: The Iliac Apophysis: An Invaluable Sign in the Management of Scoliosis. *Clinical Orthopaedics and Related Research®*. 2010;468(3):646-53.
47. Manring MM, Calhoun J. Joseph C. Risser Sr., 1892–1982. *Clinical orthopaedics and related research*. 2010;468(3):643-5.
48. Gomez JA, Hresko MT, Glotzbecker MP. Nonsurgical Management of Adolescent Idiopathic Scoliosis. *The Journal of the American Academy of Orthopaedic Surgeons*. 2016;24(8):555-64.
49. Wang WW, Xia CW, Zhu F, Zhu ZZ, Wang B, Wang SF, et al. Correlation of Risser sign, radiographs of hand and wrist with the histological grade of iliac crest apophysis in girls with adolescent idiopathic scoliosis. *Spine*. 2009;34(17):1849-54.
50. Pyle SI, Hoerr NL. A radiographic standard of reference for the growing knee: C. C. Thomas; 1955.
51. Xi HJ, Roche AF. Differences between the hand-wrist and the knee in assigned skeletal ages. *Am J Phys Anthropol*. 1990;83(1):95-102.
52. Roche AF, French NY. Differences in skeletal maturity levels between the knee and hand. *The American journal of roentgenology, radium therapy, and nuclear medicine*. 1970;109(2):307-12.
53. Urschler M, Krauskopf A, Widek T, Sorantin E, Ehammer T, Borkenstein M, et al. Applicability of Greulich-Pyle and Tanner-Whitehouse grading methods to MRI when assessing hand bone age in forensic age estimation: A pilot study. *Forensic Sci Int*. 2016;266:281-8.

54. Edmonds EW, Bathen M, Bastrom TP. Normal Parameters of the Skeletally Immature Knee: Developmental Changes on Magnetic Resonance Imaging. *Journal of pediatric orthopedics*. 2015;35(7):712-20.
55. Terada Y, Kono S, Tamada D, Uchiumi T, Kose K, Miyagi R, et al. Skeletal age assessment in children using an open compact MRI system. *Magnetic Resonance in Medicine*. 2013;69(6):1697-702.
56. Dedouit F, Auriol J, Rousseau H, Rougé D, Crubézy E, Telmon N. Age assessment by magnetic resonance imaging of the knee: A preliminary study. *Forensic Science International*. 2012;217(1):232.e1-.e7.
57. Krämer JA, Schmidt S, Jürgens K-U, Lentschig M, Schmeling A, Vieth V. The use of magnetic resonance imaging to examine ossification of the proximal tibial epiphysis for forensic age estimation in living individuals. *Forensic Science, Medicine, and Pathology*. 2014;10(3):306-13.
58. Kramer JA, Schmidt S, Jurgens KU, Lentschig M, Schmeling A, Vieth V. Forensic age estimation in living individuals using 3.0 T MRI of the distal femur. *Int J Legal Med*. 2014;128(3):509-14.
59. Fan F, Zhang K, Peng Z, Cui JH, Hu N, Deng ZH. Forensic age estimation of living persons from the knee: Comparison of MRI with radiographs. *Forensic Sci Int*. 2016;268:145-50.
60. Pennock AT, Bomar JD, Manning JD. The Creation and Validation of a Knee Bone Age Atlas Utilizing MRI. *J Bone Joint Surg Am*. 2018;100(4):e20.
61. Auf der Mauer M, Säring D, Stancus B, Herrmann J, Groth M, Jopp-van Well E. A 2-year follow-up MRI study for the evaluation of an age estimation method based on knee bone development. *International Journal of Legal Medicine*. 2018.
62. Štern D, Kainz P, Payer C, Urschler M, editors. *Multi-factorial Age Estimation from Skeletal and Dental MRI Volumes2017*; Cham: Springer International Publishing.
63. Urschler M, Grassegger S, Štern D. What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents. *Annals of Human Biology*. 2015;42(4):358-67.
64. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Aksoy S, Kazimoglu C. Forensic age estimation via 3-T magnetic resonance imaging of ossification of the proximal tibial and distal femoral epiphyses: Use of a T2-weighted fast spin-echo technique. *Forensic Sci Int*. 2016;260:102.e1-.e7.
65. Saint-Martin P, Rerolle C, Puchoux J, Dedouit F, Telmon N. Contribution of distal femur MRI to the determination of the 18-year limit in forensic age estimation. *Int J Legal Med*. 2015;129(3):619-20.
66. Schulz R, Zwiesigk P, Schiborr M, Schmidt S, Schmeling A. Ultrasound studies on the time course of clavicular ossification. *Int J Legal Med*. 2008;122(2):163-7.
67. Schmidt S, Schmeling A, Zwiesigk P, Pfeiffer H, Schulz R. Sonographic evaluation of apophyseal ossification of the iliac crest in forensic age diagnostics in living individuals. *Int J Legal Med*. 2011;125(2):271-6.
68. Giuca MR, Mazza P, Marrapese E, Cesaretti G, Calderazzi A, Carafoli D, et al. A comparison between radiographic and sonographic assessment of hand and wrist bones for the estimation of skeletal age in the child patient. *European journal of paediatric dentistry : official journal of European Academy of Paediatric Dentistry*. 2002;3(2):79-84.
69. Tanner JM, Gibbons RD. Automatic bone age measurement using computerized image analysis. *The Journal of pediatric endocrinology*. 1994;7(2):141-5.

70. Thodberg HH. An Automated Method for Determination of Bone Age. *The Journal of Clinical Endocrinology & Metabolism*. 2009;94(7):2239-44.
71. Mansourvar M, Ismail MA, Herawan T, Gopal Raj R, Abdul Kareem S, Nasaruddin FH. Automated Bone Age Assessment: Motivation, Taxonomies, and Challenges. *Computational and Mathematical Methods in Medicine*. 2013;2013:391626.
72. Gross GW, Boone JM, Bishop DM. Pediatric skeletal age: determination with neural networks. *Radiology*. 1995;195(3):689-95.
73. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully Automated Deep Learning System for Bone Age Assessment. *Journal of Digital Imaging*. 2017;30(4):427-41.
74. Stern D, Ebner T, Bischof H, Grassegger S, Ehammer T, Urschler M. Fully automatic bone age estimation from left hand MR images. *Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2014;17(Pt 2):220-7.
75. Ostojic SM. Prediction of adult height by Tanner–Whitehouse method in young Caucasian male athletes. *QJM: An International Journal of Medicine*. 2013;106(4):341-5.
76. Bayley N, Pinneau SR. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. *J Pediatr*. 1952;40(4):423-41.
77. Thodberg HH, Jenni OG, Caflisch J, Ranke MB, Martin DD. Prediction of Adult Height Based on Automated Determination of Bone Age. *The Journal of Clinical Endocrinology & Metabolism*. 2009;94(12):4868-74.
78. Martin DD, Schittenhelm J, Thodberg HH. Validation of adult height prediction based on automated bone age determination in the Paris Longitudinal Study of healthy children. *Pediatric Radiology*. 2016;46(2):263-9.
79. Stevens PM. The role of guided growth as it relates to limb lengthening. *Journal of Children's Orthopaedics*. 2016;10(6):479-86.
80. Lee SC, Shim JS, Seo SW, Lim KS, Ko KR. The accuracy of current methods in determining the timing of epiphysiodesis. *Bone & Joint Journal*. 2013;95-B(7):993-1000.
81. Sanders JO, Herring JA, Browne RH. Posterior arthrodesis and instrumentation in the immature (Risser-grade-0) spine in idiopathic scoliosis. *J Bone Joint Surg Am*. 1995;77(1):39-45.
82. Charles YP, Dimeglio A, Canavese F, Daures JP. Skeletal age assessment from the olecranon for idiopathic scoliosis at Risser grade 0. *J Bone Joint Surg Am*. 2007;89(12):2737-44.
83. Roche AF, Wainer H, Thissen D. The RWT Method for the Prediction of Adult Stature. *Pediatrics*. 1975;56(6):1026-33.
84. MOSELEY CF. A Straight Line Graph for Leg Length Discrepancies. *Clinical Orthopaedics and Related Research*®. 1978(136):33-40.
85. PALEY D, BHAVE A, HERZENBERG JE, BOWEN JR. Multiplier Method for Predicting Limb-Length Discrepancy*. *JBSJ*. 2000;82(10):1432-46.
86. Paley D, Gelman A, Shualy MB, Herzenberg JE. Multiplier method for limb-length prediction in the upper extremity. *J Hand Surg Am*. 2008;33(3):385-91.
87. Goodfellow I, Bengio Y, Courville A. *Deep Learning*: MIT Press; 2016.
88. Duda RO, Hart PE, Stork DG. *Pattern Classification*: Wiley; 2012.
89. He K, Zhang X, Ren S, Sun J, editors. *Deep Residual Learning for Image Recognition*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 27-30 June 2016.

90. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*. 2010;29(1):196-205.
91. Kelly PM, Diméglio A. Lower-limb growth: how predictable are predictions? *Journal of Children's Orthopaedics*. 2008;2(6):407-15.
92. Mansourvar M, Shamshirband S, Raj RG, Gunalan R, Mazinani I. An Automated System for Skeletal Maturity Assessment by Extreme Learning Machines. *PLoS ONE*. 2015;10(9):e0138493.
93. Schmelting A, Reisinger W, Loreck D, Vendura K, Markus W, Geserick G. Effects of ethnicity on skeletal maturation: consequences for forensic age estimations. *Int J Legal Med*. 2000;113(5):253-8.
94. Hackman L, Black S. Age Estimation from Radiographic Images of the Knee. *Journal of Forensic Sciences*. 2013;58(3):732-7.
95. Han H, Oh S, Chang CB, Kang SB. Anthropometric difference of the knee on MRI according to gender and age groups. *Surgical and radiologic anatomy : SRA*. 2016;38(2):203-11.
96. Zeiler MD, Fergus R, editors. *Visualizing and Understanding Convolutional Networks*2014; Cham: Springer International Publishing.
97. Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *CoRR*. 2013;abs/1312.6034.