

# **Ein Multitask Deep-Learning-Modell zur simultanen Erkennung, Segmentierung und Klassifizierung von Knochtumoren auf Basis von Röntgenbildern**

**Valerie Sophie Schäfer**

Vollständiger Abdruck der von der TUM School of Medicine and Health der Technischen Universität München zur Erlangung einer Doktorin der Medizin (Dr. med.) genehmigten Dissertation.

Vorsitz: Prof. Kathrin Schumann, Ph.D.

Prüfende der Dissertation:

1. Priv.-Doz. Dr. Alexandra Gersing
2. apl. Prof. Dr. Thomas Baum

Die Dissertation wurde am 27.12.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Medicine and Health am 03.07.2024 angenommen.

Diese Arbeit widme ich meiner Familie, insbesondere meinem Mann Benjamin und  
meinem Leo.

# Inhaltsverzeichnis

1	Einleitung.....	6
2	Grundlagen.....	7
2.1	Röntgenbilder zur primären Diagnostik von Knochentumoren.....	7
2.2	Knochentumore.....	7
2.2.1	Einteilung der Knochentumore nach WHO (World Health Organisation) ..	8
2.2.2	Tumorwachstum .....	9
2.2.3	Tumorlokalisierung im Skelett.....	9
2.2.4	Manifestationsalter von Knochentumoren .....	12
2.3	Einteilung nach Lodwick.....	13
2.3.1	Lokalisation im Knochen.....	13
2.3.2	Größe und Form .....	14
2.3.3	Verkalkung der Knochenmatrix .....	15
2.3.4	Muster der Osteolyse .....	15
2.3.5	Muster der Periostreaktion .....	16
2.3.6	Gradeinteilung .....	17
2.4	Künstliche Intelligenz .....	19
2.5	Künstliche Intelligenz in der Muskuloskelettalen Radiologie.....	21
3	Material und Methoden.....	23
3.1	Datensatz .....	23
3.2	Modelltraining .....	26
3.3	Modell-Architektur .....	28
3.4	Statistische Analyse .....	29
4	Ergebnisse .....	31
4.1	Patientenmerkmale und Datensätze .....	31
4.2	Modellleistung bei der internen Testreihe .....	33
4.3	Leistung des Modells in der externen Testreihe .....	34
4.4	Leistung des Modells im Vergleich mit Radiolog*innen .....	36
4.5	Modellinterpretation und Beispiele .....	39
5	Diskussion.....	44
6	Zusammenfassung.....	48

7	Literaturverzeichnis .....	49
8	Abbildungsverzeichnis.....	54
9	Tabellenverzeichnis.....	55
10	Danksagung .....	56
11	Veröffentlichungen .....	57

## Abkürzungsverzeichnis

AKZ = aneurysmatische Knochenzyste

CL = Confidence Level

CT = Computertomographie

DICOM = Digital Imaging and Communication in Medicine

DL = Deep Learning

IoU = Intersection over Union

KI = Künstliche Intelligenz

ML = Machine Learning

MRT = Magnetresonanztomographie

NOF = Nichtossifizierendes Fibrom

NOS = not otherwise specified

PNG = Portable Network Graphics

SKZ = Solitäre Knochenzyste

WHO = World Health Organisation

# 1 Einleitung

Knochtumore können sowohl bei der gezielten Röntgenuntersuchung einer schmerzhaften Region als auch als Zufallsbefund bei der Befundung von Röntgenaufnahmen des Skelettsystems erkannt werden (Erlemann, 2009). Auf Basis der Auswertungen von Stiller et al. (Stiller CA, et al., 2013) kann gefolgert werden, dass Knochtumore innerhalb Europas selten sind. Die Inzidenz der Knochtumore in Deutschland betrug im Jahr 2013 bei Frauen 400 pro 100000 Einwohner und bei Männer 460 pro 100000 Einwohner. Ferner betrug die Sterberate bei Frauen 183 pro 100000 Einwohner und bei Männer 202 pro 100000 Einwohner (Robert Koch Institute, et al., 2018). Demzufolge gehören Knochtumore zu den seltenen Krebsarten in Deutschland. Knochenmetastasen hingegen sind durchaus häufiger. Sie treten bei 50% aller Krebspatient\*innen auf (Macedo F, 2017). Knochen sind nach Lunge und Leber das am dritthäufigsten von Metastasen betroffene Organ (Mundy, 2002).

Auf Grund der niedrigen Inzidenzen ist es eine große Herausforderung eine entsprechend Fachkenntnis bei der Befundung von Röntgenbildern zu entwickeln, um entsprechende Läsionen zuverlässig erkennen und beurteilen zu können (Siegel RL, 2019). Eine frühzeitige Erkennung und präzise Diagnose sind jedoch Voraussetzung für eine angemessene und erfolgreiche Behandlung dieser Art von Tumoren (Redondo A, 2017). Neuartige Ansätze basierend auf Künstlicher Intelligenz (KI) zeigen bei der Befundung von radiologischer Bildgebung enormes Potenzial zur korrekten Befundung. Jüngste Beispiele wie von Schacky et al. (von Schacky CE, 2020) und Krogue et al. (Krogue JD, 2020) zeigen die Fähigkeit von KI bei der Unterstützung von diagnostischen Algorithmen, zum Beispiel bei der Klassifikation von Hüftgelenksarthrose oder der Identifikation von Hüftgelenksfrakturen. Beide genannten Studien überzeugten mit einer entsprechend hohen Genauigkeit und zeigen somit das Potential für die Zukunft dieser Art von KI unterstützten diagnostischen Methoden auf.

Das Ziel der vorliegenden Dissertation ist die Vorstellung eines Deep-Learning-Modells (DL), eine spezielle Untergruppe der KI. Dieses Modell soll primäre Knochtumore auf Basis von Röntgenbildern segmentieren und klassifizieren, umso eine schnelle und angemessene Weiterbehandlung zu gewährleisten.

## 2 Grundlagen

Dieses Kapitel thematisiert grundlegende Aspekte der Radiologie, welche essenziell für die Befundung und Diagnostik von Knochentumoren sind. Daher werden in Kapitel 2.1 Röntgenbilder zur primären Diagnostik von Knochentumoren, in Kapitel 2.2 Grundlagen der Knochentumore und in Kapitel 2.3 die Klassifikation nach Lodwick erläutert.

### 2.1 Röntgenbilder zur primären Diagnostik von Knochentumoren

Die Röntgendiagnostik wird als primäres bildgebendes Verfahren für die Diagnose von Knochentumoren empfohlen, da sie die Lage der Knochenläsion, das Destruktionsmuster und die Periostreaktion sichtbar machen kann (Costelloe CM, 2013) (Mintz DN, 2014). Auch wenn die Magnetresonanztomographie bei der Eingrenzung der Differentialdiagnose oder bei einer unklaren Läsion helfen kann, indem sie extraossäre Gewebebestandteile oder die Zusammensetzung des Tumors aufzeigt, gilt die Radiographie zum aktuellen Zeitpunkt weiterhin als primäre Bildgebung (Lalam R, 2017). Der Vorteil der Röntgentechnik gegenüber anderen bildgebenden Verfahren besteht darin, dass die Dichte aller Punkte in der Bildebene in einem 2D-Bild zusammengefasst wird. Die daraus resultierenden einzigartigen anatomischen Informationen ermöglichen die effiziente Bewertung von Merkmalen, die die biologische Aktivität oder die Wachstumsrate von primären Knochentumoren widerspiegeln wie z. B. Läsionsränder, periostale Reaktion, Kortikalisausdehnung, -ausdünnung und -destruktion (Costelloe CM, 2013). Diese Destruktionsmuster zeigen die biologische Aktivität von Knochenläsionen, durch die sie als aggressiv oder nicht aggressiv eingestuft werden können (Madewell JE, 1981). Des Weiteren ist die Röntgendiagnostik bis heute das am weitesten verbreitete, in der Regel schnellste und kostengünstigste medizinische Verfahren in der Detektion von Knochenläsionen (Bandyopadhyay O, 2019) (Pfeiffer D, 2020).

### 2.2 Knochentumore

Knochenläsionen können bei der klinischen Diagnose sowohl als Zufallsbefund als auch bei gezielter Untersuchung des Skeletts identifiziert werden (Balach T, 2011). Der Großteil der Läsionen wird jedoch als Zufallsbefund entdeckt, wenn Patient\*innen aus einem anderen klinischen Grund eine bildgebende Untersuchung erfahren. Daraus lässt sich die Wichtigkeit ableiten, dass Radiolog\*innen bei allen bildgebenden

Untersuchungen, die aus etwaigen medizinischen Gründen durchgeführt werden, die Knochen kritisch beurteilen müssen (Lalam R, 2017).

Knochentumore können entweder symptomlos sein oder aber Symptome wie Schwellungen, Schmerzen und Funktionsstörungen der Gliedmaßen verursachen. Diese klinischen Merkmale sind jedoch nicht besonders hilfreich, um benigne von malignen Tumoren mit einer entsprechenden Sicherheit unterscheiden zu können (Lalam R, 2017). Maligne Tumore tendieren dazu simultan in Kombination mit Schmerzen aufzutreten, jedoch können auch einige benigne Tumore wie zum Beispiel Osteoidosteome, Osteoblastome oder Chondroblastome ebenfalls mit entsprechenden Schmerzempfindungen einhergehen (Motamedi K, 2011).

Knochenläsionen können generell hinsichtlich ihres Ursprungs und ihrer Aggressivität unterschiedlich sein, weshalb es Klassifikationen der Knochentumore gibt. Im folgenden Kapitel wird die Klassifikation der Knochentumore nach der World Health Organisation vorgestellt.

### 2.2.1 Einteilung der Knochentumore nach WHO (World Health Organisation)

Generell können Knochentumore in benigne, maligne und intermediäre Knochentumore eingeteilt werden (Fletcher, 2013) (Choi JH, 2021). Die Kategorie der intermediären Knochentumore wird in zwei weitere Subgruppen unterteilt: intermediate lokal aggressive und intermediate selten metastasierende, wobei sich letztere auch lokal aggressiv verhalten (Jundt, 2018). Tabelle 1 zeigt die Einteilung der Knochentumore der WHO aus dem Jahr 2020. Die Knochenmetastasen, welche den häufigsten malignen Prozess im Knochen darstellen, sind nicht in Tabelle 1 aufgeführt.



### 2.2.2 Tumorwachstum

Das Tumorwachstum spielt für die Einteilung des Tumors eine wesentliche Rolle. Während benigne Tumore ein langsames Wachstum aufweisen, ist bei malignen Tumoren eher mit einem schnellen Wachstum zu rechnen (Lodwick, 1964). Da schnelle Wachstumsraten auf die Malignität des Knochentumors hinweisen, ist es von großer Bedeutung dieses Wachstum auf dem Röntgenbild zu erkennen (Caracciolo JT, 2016). Schnelle Wachstumsraten sind häufig mit einer schlechteren Prognose und damit mit einem geringeren Überleben der Patient\*innen assoziiert (Lodwick GS, 1980). Mit Hilfe der Klassifikation nach Lodwick kann dieses Wachstum des Knochentumors auf Röntgenbildern eingeteilt werden. Aufgrund der fundamentalen Bedeutung dieser Klassifikation wird im Kapitel 2.3 die Einteilung nach Lodwick erläutert.

### 2.2.3 Tumorlokalisation im Skelett

Für die richtige Diagnosestellung spielt als weiterer Punkt neben dem Tumorwachstum die Lokalisation des Knochentumors ebenfalls eine entscheidende Rolle. Einerseits ist die Lokalisation innerhalb des Knochens - ob in der Metaphyse, Diaphyse oder Epiphyse (Lodwick, 1964) (siehe 2.3.1) - entscheidend, andererseits spielt ebenfalls die Lokalisation im gesamten Skelettsystem, also welcher Knochen betroffen ist, eine wichtige Rolle. *Tabelle 2* zeigt eine modifizierte Einteilung der Lokalisation von Knochentumoren in verschiedenen Skelettabschnitten, nach deren Häufigkeit geordnet, nach Unni und Inwards 2010 (Unni, 2010) auf.

Ursprung	benigne	intermediär	maligne
Chondrogen	<ul style="list-style-type: none"> <li>• Subunguale Exostose</li> <li>• Bizarre parossale osteochondromatöse Proliferation</li> <li>• Periosteales Chondrom</li> <li>• Enchondrom</li> <li>• Osteochondrom</li> <li>• Chondroblastom NOS (not otherwise specified)</li> <li>• Chondromyxoidfibrom</li> <li>• Osteochondromyxom</li> </ul>	<ul style="list-style-type: none"> <li>• Chondromatose NOS</li> <li>• Atypischer kartilaginärer Tumor</li> </ul>	<ul style="list-style-type: none"> <li>• Chondrosarkom Grad 1, 2 und 3</li> <li>• Periostales Chondrosarkom</li> <li>• Klarzelliges Chondrosarkom</li> <li>• Mesenchymales Chondrosarkom</li> <li>• Enddifferenziertes Chondrosarkom</li> </ul>
Osteogen	<ul style="list-style-type: none"> <li>• Osteom NOS</li> <li>• Osteoidosteom NOS</li> </ul>	<ul style="list-style-type: none"> <li>• Osteoblastom NOS</li> </ul>	<ul style="list-style-type: none"> <li>• Osteosarkome *</li> </ul>
Fibrös		<ul style="list-style-type: none"> <li>• Desmoplastisches Fibrom</li> </ul>	<ul style="list-style-type: none"> <li>• Fibrosarkom</li> </ul>
Osteoklastische Riesenzelltumore	<ul style="list-style-type: none"> <li>• Aneurysmatische Knochenzyste</li> <li>• Nicht-ossifiziertes Fibrom</li> </ul>	<ul style="list-style-type: none"> <li>• Riesenzelltumor</li> </ul>	<ul style="list-style-type: none"> <li>• Maligner Riesenzelltumor</li> </ul>
Notochordal	<ul style="list-style-type: none"> <li>• Benigner notochordaler Tumor</li> </ul>		<ul style="list-style-type: none"> <li>• Chordom NOS</li> <li>• Chondroid Chordom</li> <li>• Schlecht differenziertes Chordom</li> <li>• Enddifferenziertes Chordom</li> </ul>
Vaskulär	<ul style="list-style-type: none"> <li>• Hämangiom+</li> </ul>	<ul style="list-style-type: none"> <li>• Epitheloides Hämangiom</li> </ul>	<ul style="list-style-type: none"> <li>• Epitheloides Hämangioendotheliom NOS</li> <li>• Angiosarkom</li> </ul>
Andere Mesenchymale Knochentumore	<ul style="list-style-type: none"> <li>• Chondromesenchymales Hamartom der Brustwand</li> <li>• Einfache Knochenzyste</li> <li>• Fibröse Dysplasie</li> <li>• Osteofibröse Dysplasie</li> <li>• Lipom NOS</li> <li>• Hibernom</li> </ul>	<ul style="list-style-type: none"> <li>• Osteofibröse Dysplasie</li> <li>• Mesenchymom NOS</li> </ul>	<ul style="list-style-type: none"> <li>• Adamantinom der langen Knochen</li> <li>• Enddifferenziertes Adamantinom</li> <li>• Leiomyosarcom NOS</li> <li>• Pleomorphes Sarkom</li> <li>• Undifferenzierte Knochenmetastase</li> </ul>
Hämatopoetische Neoplasien des Knochens	<ul style="list-style-type: none"> <li>• Knochenplasmozytom</li> <li>• Malignes Lymphom</li> <li>• Non-Hodgkin NOS</li> <li>• Hodgkin NOS</li> <li>• Diffuses B-Zell- Lymphom NOS</li> <li>• Follikuläres Lymphom NOS</li> <li>• Marginales B-Zell-Lymphom NOS</li> </ul>		<ul style="list-style-type: none"> <li>• T-Zell-Lymphom NOS</li> <li>• Anaplastisches Großzellymphom NOS</li> <li>• Malignes Lymphom NOS</li> <li>• Burkitt Lymphom</li> <li>• Langerhans-Zell-Histozytose NOS</li> <li>• Erdheim-Chester- Erkrankung</li> <li>• Rosai-Dorfman-Erkrankung</li> </ul>

\* zu den Osteosarkomen zählen: Konventionelles Osteosarkom, Teleangiektatisches Osteosarkom, Kleinzelliges Osteosarkom, Periostales Osteosarkom, Parossales Osteosarkom, Hochmalignes oberflächliches Osteosarkom, Sekundäres Osteosarkom, Niedrigmalignes zentrales Osteosarkom

Tabelle 1: Einteilung der Knochentumore nach WHO 2020 in Anlehnung an (Choi JH, 2021).

Lokalisation	Benigne und intermediäre Tumore	Maligne Tumore
Becken	Riesenzelltumor, Osteoblastom, Neurilemmom (Schwannom)	Chordom, Chondrosarkom, Osteosarkom, Ewing-Sarkom
Femur	Osteochondrom, NOF, Riesenzelltumor, Osteoidosteom, Chondrom	Osteosarkom, Chondrosarkom, Ewing-Sarkom, Myelom
Fibula	Osteochondrom, Riesenzelltumor, Chondrom, NOF	Osteosarkom, Ewing-Sarkom, Chondrosarkom
Humerus	Osteochondrom, Chondrom, Riesenzelltumor, SKZ, AKZ	Osteosarkom, Chondrosarkom, Myelom, Ewing-Sarkom
Karpus	Osteoidosteom, Riesenzelltumor, Chondrom, SKZ	Chondrosarkom
Metakarpalia	Chondrom, Osteochondrom, Osteoidosteom	Chondrosarkom, Osteosarkom
Metatarsus	Chondrom, Osteoidosteom, Osteochondrom	Ewing-Sarkom, Chondrosarkom, Osteosarkom
Patella	Chondroblastom, SKZ, AKZ	Lymphom
Radius	Riesenzelltumor, Osteochondrom, Osteoidosteom, Chondrom	Osteosarkom, Ewing-Sarkom, Chondrosarkom
Rippen	Fibröse Dysplasie, Osteochondrom, Riesenzelltumor	Myelom, Chondrosarkom, Ewing-Sarkom
Schädel	Hämangiom, Eosinophiles Granulom, fibröse Dysplasie	Chordom, Osteosarkom, Myelom
Skapula	Osteochondrom, Chondrom, Chondroblastom, Osteoidosteom	Chondrosarkom, Osteosarkom, Myelom, Ewing-Sarkom
Sternum	Chondrom, Riesenzelltumor	Myelom, Chondrosarkom
Tarsus	Osteoidosteom, Chondroblastom, Osteochondrom, SKZ	Osteosarkom, Ewing-Sarkom, Chondrosarkom
Tibia	Riesenzelltumor, Osteochondrom, Osteoidosteom, Chondroblastom	Osteosarkom, Chondrosarkom, Ewing-Sarkom, Fibrosarkom, Adamantinom
Ulna	Riesenzelltumor, Osteoidosteom, Osteochondrom	Ewing-Sarkom, Osteosarkom
Wirbelkörper	Riesenzelltumor, Hämangiom, Osteoidosteom, Osteoblastom	Myelom, Chordom, Chondrosarkom, Osteosarkom

AKZ: aneurysmatische Knochenzyste  
SKZ: solitäre Knochenzyste

Tabelle 2: Lokalisation der Knochentumore im Skelett, in Anlehnung an (Unni, 2010).

## 2.2.4 Manifestationsalter von Knochentumoren

Das Alter der Patient\*innen ist bei der Diagnose von Knochentumoren ein fundamentaler Faktor. Die Entstehung von benignen Knochentumoren nach dem 40. Lebensjahr ist selten. Demzufolge sollte jede neu diagnostizierte Knochenläsion bei Patient\*innen über 40 Jahren bis zum Beweis des Gegenteils als maligner Tumor klassifiziert werden. Eine Ausnahme stellen jedoch Knochenläsionen in der Nähe von Gelenken dar, da diese z.B. auch eine subchondrale Zyste im Zusammenhang mit Osteoarthritis oder einer anderen Gelenkerkrankung sein können. Bei Patient\*innen unter 40 Jahren hingegen kann der Knochentumor entweder benigne oder maligne sein. Das Patient\*innenalter stellt einen zuverlässigen Faktor zur Erstellung einer Differentialdiagnose dar und muss bei der Befundung des Röntgenbilds mitbeurteilt werden. *Tabelle 3* zeigt das typische Manifestationsalter von unterschiedlichen Knochentumoren (Lalam R, 2017).

Alter in Jahren	aggressiv	nicht aggressiv
0-10	<ul style="list-style-type: none"> <li>• Ewing-Sarkom</li> <li>• eosinophiles Granulom,</li> <li>• Neuroblastom,</li> <li>• Hämatologische Malignome, die den Knochen betreffen</li> </ul>	<ul style="list-style-type: none"> <li>• Einfach Knochenzyste</li> <li>• eosinophiles Granulom</li> <li>• aneurysmatische Knochenzyste</li> </ul>
10-20	<ul style="list-style-type: none"> <li>• Osteosarkom</li> <li>• Ewing-Sarkom</li> <li>• Adamantinom</li> <li>• Riesenzelltumor</li> </ul>	<ul style="list-style-type: none"> <li>• Nichtossifizierendes Fibrom</li> <li>• Fibröse Dysplasie</li> <li>• einfache Knochenzyste</li> <li>• aneurysmatische Knochenzyste</li> <li>• Osteochondrom</li> <li>• Chondroblastom</li> <li>• Chondromyxoidfibrom</li> <li>• Adamantinom</li> </ul>
20-40	<ul style="list-style-type: none"> <li>• Chondrosarkom</li> <li>• parossales Osteosarkom</li> <li>• pleomorphes Sarkom</li> </ul>	<ul style="list-style-type: none"> <li>• Enchondrom</li> <li>• Riesenzelltumor</li> </ul>
> 40	<ul style="list-style-type: none"> <li>• Metastasen</li> <li>• Multiples Myelom</li> <li>• Plasmozytom</li> <li>• Chondrosarkom</li> <li>• Osteosarkom</li> <li>• pleomorphes Sarkom</li> <li>• Chordom</li> <li>• Lymphom</li> </ul>	<ul style="list-style-type: none"> <li>• Interossäres Ganglion</li> <li>• Subchondrale Zyste</li> </ul>

Tabelle 3: Typisches Manifestationsalter von Knochentumoren, in Anlehnung an (Lalam R, 2017).

## 2.3 Einteilung nach Lodwick

Gwilym S. Lodwick stellt in seinen Artikeln („*Radiographic Diagnosis and Grading of Bone Tumors, With Comments on Computer Evaluation*“ aus dem Jahr 1964 und in einer überarbeiteten Version aus dem Jahr 1980 „*Determining growth rates of focal lesions of bone from radiographs*“) ein Klassifikationssystem vor (Lodwick, 1964) (Lodwick GS, 1980). Primäres Ziel dieser Gradeinteilung ist es Knochentumore auf konventionellen Röntgenbildern in einen von drei Graden einzuteilen, welche die Unterschiede in der Wachstumsgeschwindigkeit und damit auch der Aggressivität widerspiegeln. Diese Veränderung im Knochen geschieht durch osteoklastäre und/oder osteoblastäre Reaktionen, welche die Knochenstruktur verändern. Nach Lodwick sind zum einen die Lokalisation, Größe und Form zur Diagnose des Tumors von großer Bedeutung und zum anderen die Knochendestruktion bzw. Proliferation sowie die Verkalkung der Tumormatrix. All diese bildmorphologischen Faktoren müssen zusätzlich mit dem Alter der Patient\*innen in Zusammenhang gebracht werden, um eine Diagnose erstellen zu können. Im Folgenden werden detailliert die unterschiedlichen Kriterien zur Diagnosestellung nach Lodwick vorgestellt.

### 2.3.1 Lokalisation im Knochen

Knochentumore kommen an spezifischen Orten des Skeletts vor (vgl. Kapitel 2.2.3). In der Regel treten Primärtumore in den Röhrenknochen, insbesondere am distalen Femur, proximale Tibia sowie proximaler Humerus auf, wo das stärkste Wachstum lokalisiert ist. Auch innerhalb der langen Röhrenknochen treten die Knochentumore an spezifisch bevorzugten Abschnitten auf. So kommen beispielsweise Riesenzelltumore und Chondroblastome vorwiegend in der Epiphyse vor. Chondrosarkome und Osteosarkome befinden sich bevorzugt in der Metaphyse und die typische Lokalisation von Ewing-Sarkomen ist in der Diaphyse (Miller, 2008). Die nachfolgende Abbildung 1 zeigt schematisch typische Lokalisationen von Knochentumoren bezogen auf das Vorkommen im Knochen.

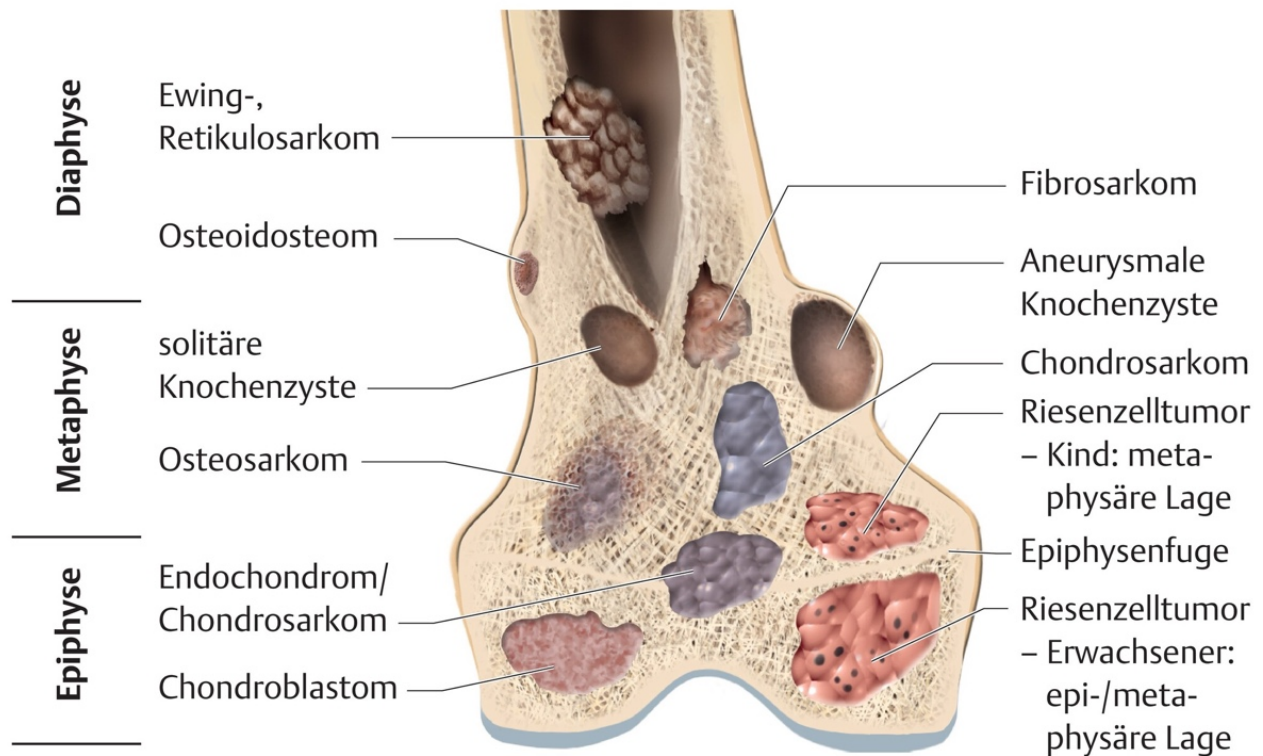


Abbildung 1: Typische topografische Lokalisation von Knochentumoren (Uhl M, 2015).

### 2.3.2 Größe und Form

Langsam wachsende Tumore sind bei Entdeckung meist unter 6cm im Durchmesser groß. Demgegenüber stehen maligne Tumore, die meist über 6cm im Durchmesser aufweisen. Daraus lässt sich ableiten, dass die meisten Primärtumore, die zum Zeitpunkt der Erstentdeckung klein sind, einen niedrigen Malignitätsgrad aufweisen. Die meisten primären Knochentumore wachsen progressiv in alle Richtungen und neigen dabei dazu, eine kugelförmige Gestalt anzunehmen. Eine Ausnahme stellt hier das Ewings-Sarkom dar, welches bei Lokalisation in der Diaphyse meist länglich ausgeprägt ist (Lodwick, 1964).

### 2.3.3 Verkalkung der Knochenmatrix

Knochenbildende Tumore, die eine funktionelle Komponente beibehalten, weisen im Röntgenbild häufig Bereiche mit erhöhter Dichte auf, welche die Bildung von Tumorknochen aufzeigen. Dieser Tumorknochen entsteht, wenn das Tumorgewebe Osteoid oder eine chondrogene Matrix produziert, welches in einem nächsten Schritt mineralisiert wird. In Abhängigkeit von der Tumormineralisation wird auf dem Röntgenbild das Ausmaß der Dichteanhebung unterschieden (Erlemann, 2009). Hierbei wird nach einer sehr dichten elfenbeinartigen Knochenverdichtung, nach einer wolkenartigen oder einer erkennbaren diffusen Knochenverdichtung differenziert (Uhl M, 2022).

### 2.3.4 Muster der Osteolyse

Die Destruktion des Knochens als Folge des Tumorwachstums zeigt mit einer hohen Wahrscheinlichkeit eines der drei folgenden röntgenologischen Muster:

#### 1. Geografische Osteolyse

Bei diesem Muster erzeugt der Tumor eine einzelne große, gut definierte Osteolyse oder mehrere zusammenhängende Osteolysen im spongiosen oder kortikalen Knochen. Diese Osteolysen sind häufig überlappend und haben scharf begrenzte Ränder. Im Frühstadium dieser Form der Knochendestruktion kann nur eine schwache Transparenz der Kortikalis sichtbar sein. Dagegen kann es im Endstadium zu einer pathologischen Knochenfraktur kommen oder der gesamte Knochen kann destruiert sein (Lodwick, 1964).

#### 2. Mottenfraßartige Destruktion des Knochens

Einige Tumore erzeugen ein Muster aus vielen verstreuten und konfluierenden Osteolysen von mittlerer Größe, welche sich um mehrere, räumlich voneinander getrennte, Herde anordnen. Die mottenzerfressene Destruktion ist in der Regel inhomogen und verstreut. Im kortikalen Knochen führt diese Art der Destruktion zu einer pathologischen Fraktur (Lodwick, 1964).

#### 3. Permeative Osteolyse

Das dritte Hauptmuster der Knochendestruktion ist gekennzeichnet durch zahlreiche winzige Osteolysen in der Kortikalis, die in Größe und Häufigkeit unmerklich abnehmen, von der Zone des maximalen Befalls bis zur Kortikalis, welche intakt ist. Eine klare Begrenzung ist bei diesem Muster der Knochendestruktion nicht sichtbar. Bei diesem unregelmäßigen permeativen Muster ist es weniger wahrscheinlich, dass es zu einer frühen pathologischen Fraktur kommt. Eine frühe permeative Destruktion kann auf dem Röntgenbild jedoch unsichtbar sein (Lodwick, 1964).

Zusammenfassend zeigt *Abbildung 2* eine schematische Darstellung der möglichen Knochenosteolysen.

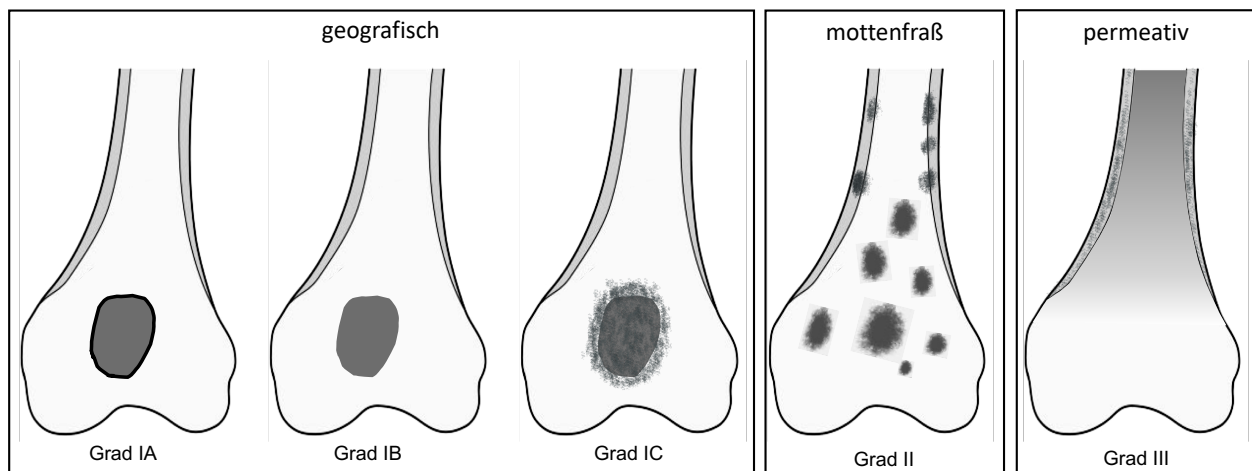


Abbildung 2: Schematische Darstellung der Knochenosteolysen, in Anlehnung an (Lalam R, 2017)

### 2.3.5 Muster der Periostreaktion

Ein weiterer Hinweis auf die Wachstumsgeschwindigkeit einer Läsion ist die Periostreaktion. Eine glatte und kontinuierliche periostale Knochenneubildung ist das Zeichen für einen nicht aggressiven Tumor, da das Periost durch das langsame Wachstum des Tumors genügend Zeit hat, eine kontinuierliche Schicht neuen Knochens zu bilden (Ragsdale BD, 1981). Einzelne Periostlamellen oder nicht verkalkte Periostschalen weisen auf ein intermediäres Wachstum hin.

Aggressive Tumore können allerdings durch ihr schnelles Wachstum eine ungeordnete, diskontinuierliche und komplexe Periostreaktion hervorrufen. Diese werden unter anderem als Codman-Dreieck, Spikula oder Strebepfeilerform beschrieben. Eine dreieckige Abhebung des Periosts, welche zu aggressiven Läsionen benachbart ist, die zu einer Durchdringung der Kortikalis geführt hat, nennt man Codman-Dreieck. Knochenneubildungen mit dazwischenstehenden Gefäßneubildungen bezeichnet man als Spikulae, wenn diese senkrecht und parallel zur Kortikalis angeordnet sind. Oberflächentumore werden durch eine Strebepfeilerform charakterisiert. *Abbildung 3* zeigt eine schematische Darstellung der möglichen Periostreaktionen.



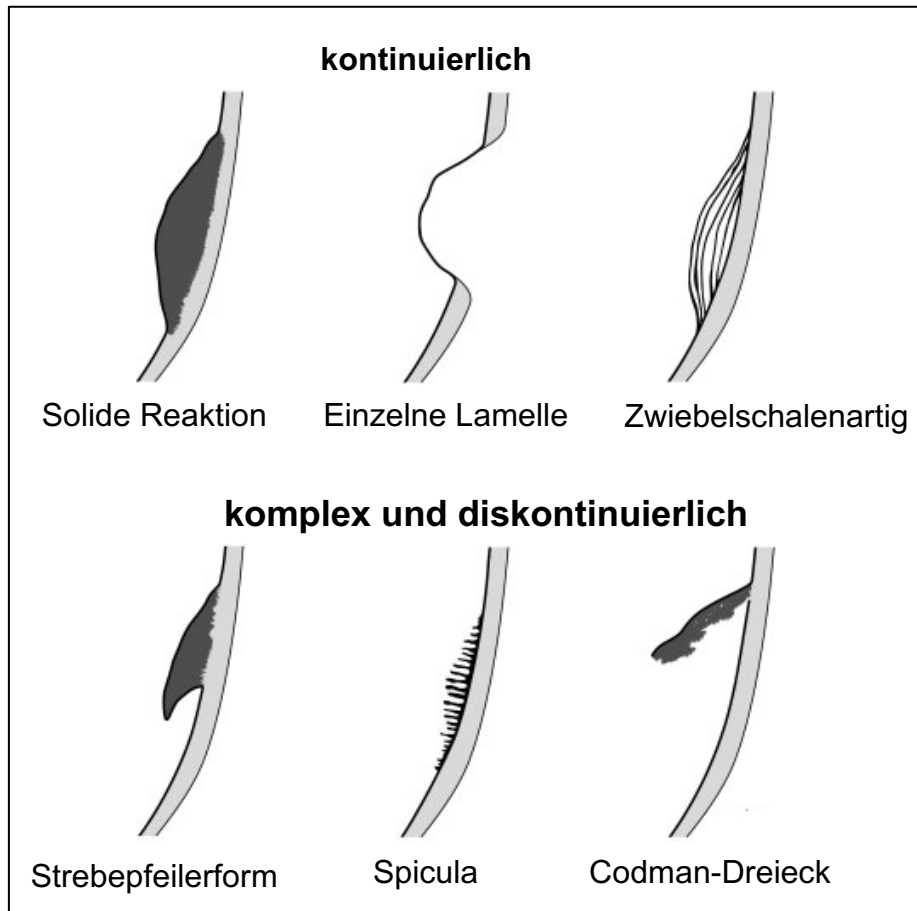


Abbildung 3: Übersicht der Periostreaktionen, in Anlehnung an (Lalam R, 2017)

### 2.3.6 Gradeinteilung

Ziel der Klassifizierung von Lodwick ist es, den Knochentumor nach seiner Wachstumsgeschwindigkeit einzuteilen. Um diese Klassifizierung vornehmen zu können, müssen alle oben genannten Punkte in Bezug zueinander gesetzt werden. Lodwick nennt in seiner Studie 18 Punkte, die bei der Einteilung unterstützen sollen. Nahezu alle Knochentumore können in eine von fünf Kategorien (IA, IB, IC, II, III) eingeteilt werden, die jeweils einen unterschiedlichen Bereich der Wachstumsrate darstellen. *Tabelle 4* zeigt eine tabellarische Übersicht der Einteilung nach Lodwick, in welcher die Lage und Begrenzung der Knochenosteolyse, die Penetration der Kortikalis sowie auch der Sklerosesaum samt Expansion beschrieben sind. Zur Visualisierung ist ein Schema der Knochenosteolysen zusätzlich am Tabellenende von *Tabelle 4* angefügt. Zudem ist in *Abbildung 4* ein Entscheidungsbaum dargestellt, welcher bei der Befundung von Knochenläsionen eine Hilfestellung leisten soll.

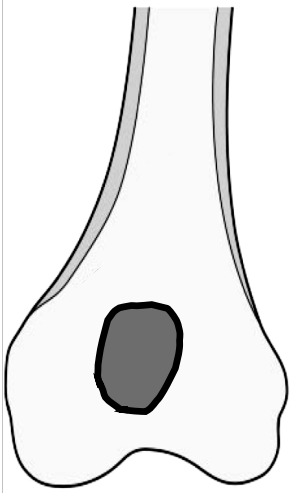
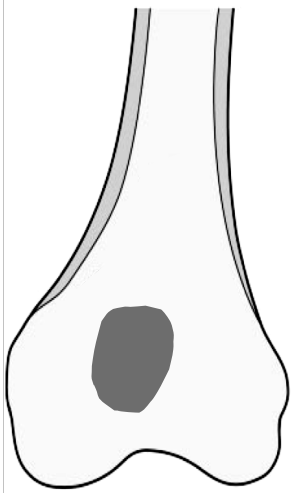

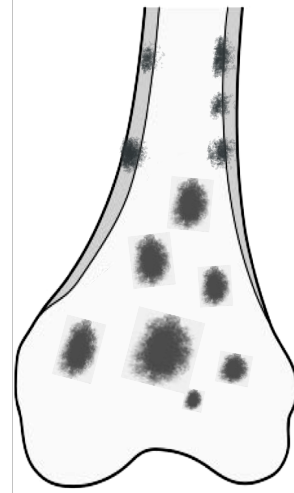

Röntgenmuster	IA	IB	IC	II	III
<b>Knochenosteolyse</b>	geografisch	geografisch	geografisch	geografisch + mottenfraß und/oder permeativ	mottenfraß und/oder permeativ
<b>Begrenzung</b>	regelmäßig oder lobuliert, immer scharf	regelmäßig oder lobuliert, auch multizentrisch	regelmäßig oder lobuliert, auch multizentrisch, oder unregelmäßig/unscharf, aber nicht mottenfraß	unscharf	unscharf
<b>Penetration der Kortikalis</b>	nicht oder teilweise	nicht oder teilweise	vollständig	vollständig	vollständig
<b>Sklerosesaum</b>	immer	möglich	möglich	möglich, aber ungewöhnlich	möglich, aber ungewöhnlich
<b>Expansion der Sklerose</b>	möglich aber $\leq 1\text{cm}$	$> 1\text{cm}$ , Sklerosesaum vorhanden	möglich	möglich, aber ungewöhnlich	möglich, aber ungewöhnlich
<b>Schematische Darstellung</b>					

Tabelle 4: Übersicht der Einteilung nach Lodwick Grad I-III (Lodwick GS, 1980)

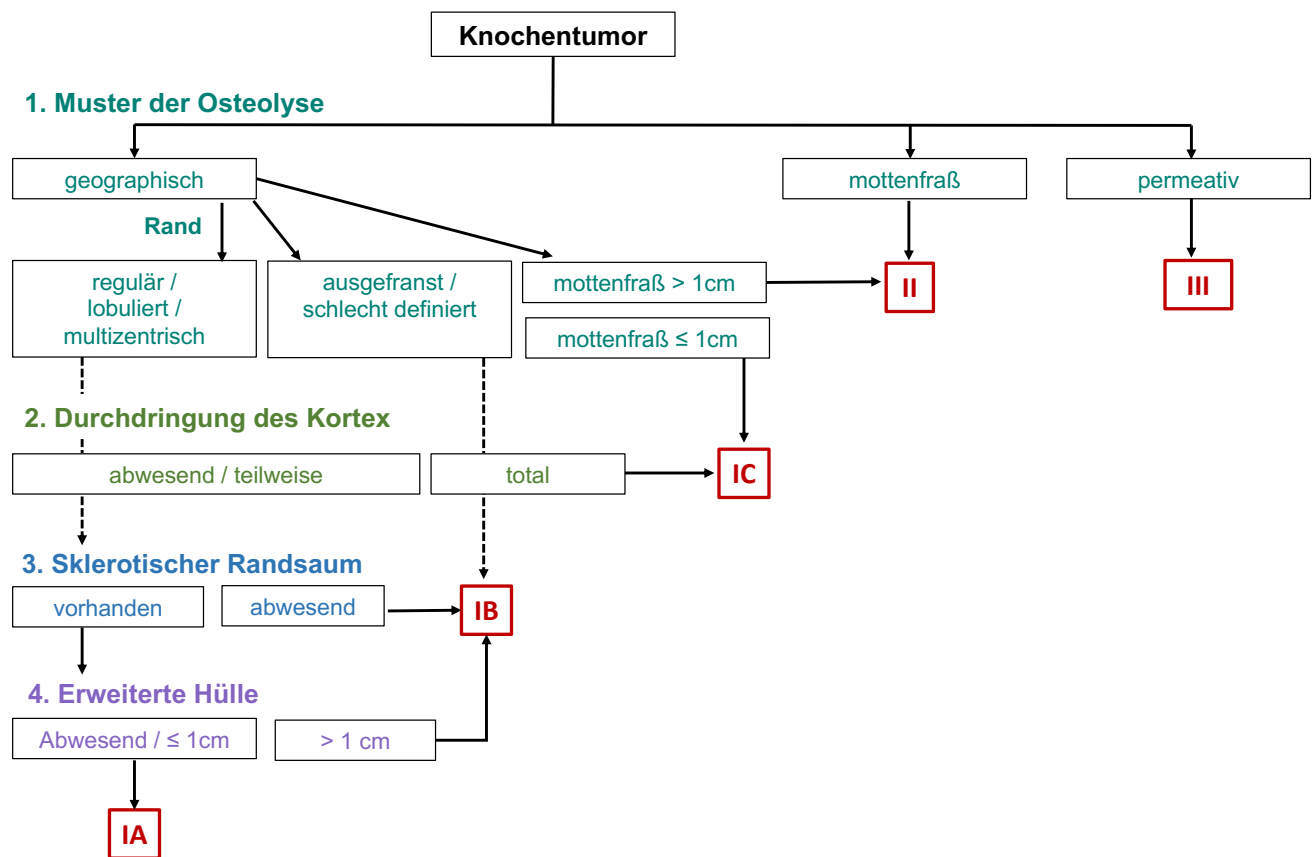


Abbildung 4: Entscheidungsbaum zur Klassifizierung nach Lodwick Grad I-III (Lodwick GS, 1980)

## 2.4 Künstliche Intelligenz

Einer der vielversprechendsten Bereiche der Gesundheitsinnovation ist die Anwendung von KI in der medizinischen Bildgebung (Lakhani P, 2018). KI wurde in den 1950er Jahren erstmals von McCarthy eingeführt und ist ein allgemeiner Begriff, der Computermaschinen beschreibt, die die menschliche Intelligenz imitieren (Han XG, 2019). Dabei bezieht sich KI auf jene Technik, die es Computern ermöglicht, die menschliche Intelligenz nachzuahmen (Poole DI, 1998). Der Bereich KI enthält unter anderem zwei weitere Subkategorien: Machine Learning (ML) und Deep Learning (DL), wie in *Abbildung 5* schematisch dargestellt (Gyftopoulos S, 2019). ML ist ein spezielles Teilgebiet der KI, welches es Maschinen ermöglicht, die Leistung von Aufgaben mit Hilfe von verschiedenen Werkzeugen aus der Statistik, Mathematik und Informatik zu verbessern (Gyftopoulos S, 2019). ML kann durch Algorithmen selbständig Muster in Daten erkennen. Entscheidend ist jedoch, dass die Daten zuvor vorbereitet und strukturiert werden müssen, bevor sie verwendet werden können. Deep Learning hingegen zeichnet sich durch die selbstständige Vorbereitung und Strukturierung der Daten aus (Chauhan NK, 2018). Die Entstehung von DL ist nur durch die rasanten Fortschritte der Rechenleistungen und die Verfügbarkeit großer Datenmengen möglich (Vogrin M, 2020). DL verwendet sogenannte künstliche neuronale Netze, für welche großen Datensätzen Voraussetzung sind (LeCun Y, 2015).

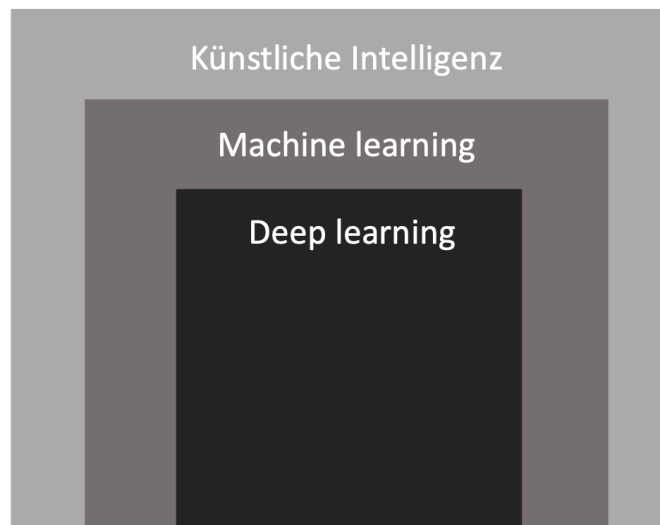


Abbildung 5: Künstliche Intelligenz und die Unterkategorien Machine learning und Deep learning

Diese künstlichen neuronale Netze, welche Berechnungsmodelle und Algorithmen umfassen, ähneln der Struktur und Funktion des menschlichen Gehirns (King, 2017). Sie sind in der Lage zu lernen, indem sie Informationen in Form von digitalen Daten erhalten, verarbeiten und ihre Struktur in ähnlicher Weise verändern wie die Neuronen in unserem Gehirn verändert und verschaltet werden (King, 2017). Die künstlichen Neuronen sind in sogenannten neuronalen Netzwerken organisiert. Diese neuronalen Netzwerke sind in Schichten strukturiert, die aus miteinander verbundenen Knotenpunkten bestehen. Diese Schichten werden in drei verschiedene Arten aufgeteilt. Die Eingabeschicht, welche die Eingabedaten empfängt, die verborgene Schicht, welche die Muster in den Daten extrahiert und die Ausgabeschicht, welche anschließend die Ergebnisse der Datenverarbeitung liefert (Erickson BJ, 2017). Diese unterschiedlichen Schichten, visualisiert in *Abbildung 6*, bestehen aus einer Vielzahl von künstlichen Neuronen, die auf unterschiedliche Art und Weise miteinander verbunden sind. Die Neuronen erhalten als Eingabe eine Vielzahl von Werten, welche Merkmale oder Attribute darstellen. Diese werden jeweils mit einer entsprechenden Gewichtung im Netzknoten multipliziert. Die nun gewichteten Merkmale oder Attribute werden im Anschluss summiert und durch eine nichtlineare Aktivierungsfunktion geleitet. Die Gewichtung der Eingangsdaten wird während der Trainingsphase dynamisch optimiert und angepasst. Ein künstliches Neuron kann demnach so schematisch dargestellt werden, dass es eine Entscheidung auf Basis von Abwägen einer Reihe von Beweisen trifft (Chartrand G, 2012).

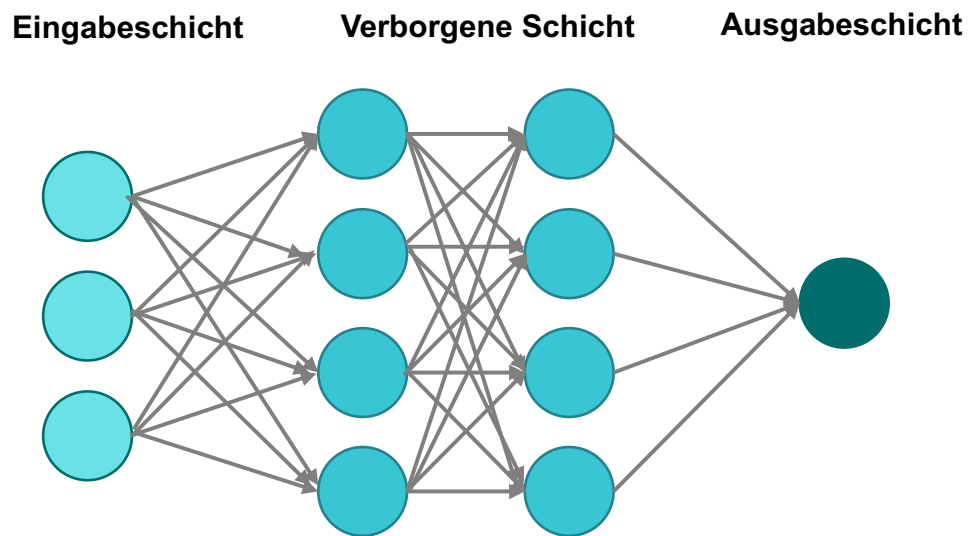


Abbildung 6: Visualisierung der drei Schichten der neuronalen Netzwerke

## 2.5 Künstliche Intelligenz in der Muskuloskelettalen Radiologie

In den letzten Jahren wurden immer mehr Studien in der Muskuloskelettalen Radiologie zum Thema KI durchgeführt. So haben Bayramoglu et al. in ihrer Studie zur Detektion von Arthrose im Kniegelenk herausgefunden, dass der Einsatz von KI (Radiomics) bei der Erkennung des Vorhandenseins von Arthrose des Kniegelenks den bisherigen Ansätzen zur Erkennung von Arthrose auf Röntgenbildern überlegen ist (Bayramoglu N, 2020). In einer weiteren Studie zur Osteoarthrose des Kniegelenks prognostizierte das Deep-Learning-Modell das Risiko eine Kniegelenktotalendoprothese zu erhalten besser voraus als bisherige Standard-Grading-Systeme (Leung K, 2020). Von Schacky et al. haben in ihrer Studie zur Beurteilung von radiologischen Merkmalen der Hüftarthrose auf Röntgenbildern demonstriert und aufgezeigt, dass der Multitask-Deep-Learning-Ansatz mit der Leistung von Radiolog\*innen vergleichbar ist (von Schacky CE, 2020). Eine große Anzahl von Studien zum Einsatz von KI gibt es auch aus dem Bereich der Detektion von Frakturen. Adams et al. zeigen beispielsweise in ihrer Studie, dass ein Deep-Learning-Modell in der Detektion von Schenkelhalsfrakturen die gleichen Ergebnisse erzielt wie ein\*e Radiolog\*in (Adams M, 2019). Eine weitere Studie zeigt, dass der Einsatz von KI proximale Humerusfrakturen auf einfachen AP-Röntgenaufnahmen der Schulter genau erkennen und klassifizieren kann (Chung SW, 2018). Auch aus dem Bereich der Knochentumordiagnostik gibt es bereits vielversprechende Studien. Bandyopadhyay et al. haben eine Technik für die automatische Diagnose von Knochenkrebs in langen Röhrenknochen vorgestellt, die ausschließlich auf der Analyse eines Röntgenbildes basiert. Dabei stimmen in 85 % der Fälle das Muster der Knochendestruktion, das Stadium und der Grad des Knochentumors, die von der KI vorhergesagt wurden, mit den tatsächlichen Befunden überein (Bandyopadhyay O, 2019).

In einer weiteren Studie wurde kürzlich DL zur Klassifizierung von primären Knochentumoren auf Röntgenbildern verwendet, jedoch wurden in dieser Studie die Bilder manuell um den Tumor extrahiert, was die klinische Anwendbarkeit stark einschränkt (He Y, 2020).

## 3 Material und Methoden

### 3.1 Datensatz

Die hier beschriebene Studie ist eine multizentrische Studie, welche von den lokalen institutionellen Ethikkommissionen genehmigt wurde (von Schacky CE, 2021). Die Studie erfolgte unter Einhaltung der nationalen und internationalen Richtlinien.

In die Studie eingeschlossen sind alle Patient\*innen mit Röntgenbildern von Knochtumoren (2003 bis 2020), die im Interdisziplinären Muskuloskelettalen Tumorzentrum des Klinikums der Technischen Universität München aufgenommen worden waren, keine vorhergegangene Therapie wie Operation oder Chemotherapie erhalten hatten und bei denen ein histopathologischer Befund der knöchernen Läsion vorlag. Dieser histologische Befund wurde als Referenzstandard für die eigentliche Diagnose der Knochtumore auf den Röntgenbildern verwendet.

Die Röntgenbilder stammen aus der Patientendatenbank des Interdisziplinären Muskuloskelettalen Tumorzentrums des Klinikums der Technischen Universität München. Die Bilddaten wurden aus Digital Imaging and Communication in Medicine (DICOM-Dateien) als PNG-Dateien (Portable Network Graphics) extrahiert. PNG wurde gewählt, um eine weitere, verlustfreie Bildverarbeitung zu gewährleisten.

In der Studie sind Patient\*innen berücksichtigt worden, bei denen entweder die Diagnose eines benignen oder eines malignen Tumors vorlag. Hierbei zählen zu den benignen Tumoren Osteochondrome, Enchondrome, Chondroblastome, Osteoidosteome, nichtossifizierende Fibrome, Hämangiome, Riesenzelltumoren, einfache Knochenzysten, aneurysmatische Knochenzysten und fibröse Dysplasie. Zu den malignen Tumoren werden Chondrosarkome, Osteosarkome, Ewing-Sarkome, Plasmazellmyelome, B-Zell-Non-Hodgkin-Lymphom und Chordome gezählt. Als Referenzstandard gilt die histologische Diagnose, die anhand einer Biopsie oder Gewebeprobe nach einer Resektion gestellt wurde. Ein weiteres Einschlusskriterium für die Studie war es, dass auf dem Röntgenbild nur ein einziger Knochtumor zu sehen ist und keine multiplen Läsionen. So war es möglich, im primären Schritt 1132 Patient\*innen in die Studie mit aufzunehmen.

In einem zweiten Schritt führten zwei Radiolog\*innen (A.S.G., mit 8 Jahren Erfahrung; S.C.F., mit 4 Jahren Erfahrung) eine randomisierte Überprüfung der 1132 ausgewählten Röntgenbildern als unabhängige Reader durch. Die Radiolog\*innen hatten zum Zeitpunkt der Readings dabei keinen Zugang zu Patienteninformationen oder histopathologischen Befunden. In diesem zweiten Schritt wurden Röntgenbilder ausgeschlossen, bei denen die Bildqualität als nicht ausreichend durch die Radiolog\*innen klassifiziert wurde, zum Beispiel aufgrund von Artefakten oder weil der Tumor nicht auf dem Röntgenbild abgebildet war. Wenn sich beide Radiolog\*innen einig waren, dass keine Segmentierung und sinnvolle Analyse anhand des Röntgenbildes durchgeführt werden konnten, galt die Bildqualität als nicht ausreichend. Dadurch kam es zum Ausschluss von weiteren 198 Röntgenbildern.

Außerdem wurden 72 digitalisierte Filmröntgenbilder von 72 Patient\*innen mitberücksichtigt. *Abbildung 7* zeigt ein Flussdiagramm, in welchem die Patientenauswahl dargestellt ist. Der interne Datensatz beinhaltet nach allen Ausschlüssen 934 Röntgenbilder, die für die Studie zur Verfügung stehen.

Die Röntgenbilder (n = 934) wurden mittels Split-Sample-Validierung in Trainings-, Validierungs- und Testgruppen aufgeteilt. Gemäß der Split-Sample-Validierung wurden 70 % der Gesamtstichprobe (n = 654) in die Trainingsgruppe, 15 % (n = 140) in die Validierungsgruppe und 15 % (n = 140) in die Testgruppe nach dem Zufallsprinzip eingeordnet, um die Anzahl der Tumortypen in allen drei Gruppen auszugleichen. Darüber hinaus stand ein externer Testsatz von einem weiteren externen Universitätsklinikum zur Verfügung, welcher im gleichen Datenformat vorlag. Dieser Datensatz umfasst 111 Röntgenbilder von Patient\*innen mit benignen Tumoren (einschließlich Osteochondrom, Enchondrom, Chondroblastom, osteoid Osteom, nichtossifizierendes Fibrom, Hämangiom, Riesenzelltumor, einfache Knochenzyste, aneurysmatische Knochenzyste und fibröse Dysplasie) und malignen Tumoren (einschließlich Chondrosarkom, Osteosarkom, Ewing-Sarkom, Plasmazell-Myelom und B-Zell-NHLs). Zweck dieses externen Testsatzes, welcher die Zielpopulation adäquat repräsentiert, ist die geografische Validierung, welche besonders wichtig ist, um eine Überschätzung der Ergebnisse aufgrund von Overfitting des Modells zu vermeiden (Park SH, 2018). Overfitting bezieht sich auf eine Situation, in der sich ein Lernmodell zu sehr an die Trainingsdaten anpasst und zwar in einem Maße, das es sich negativ auf die Fähigkeit des Modells auswirkt, neue Daten zu verallgemeinern, während die Leistung des Modells mit dem Trainingsdatensatz übertrieben wird (Park SH, 2018).



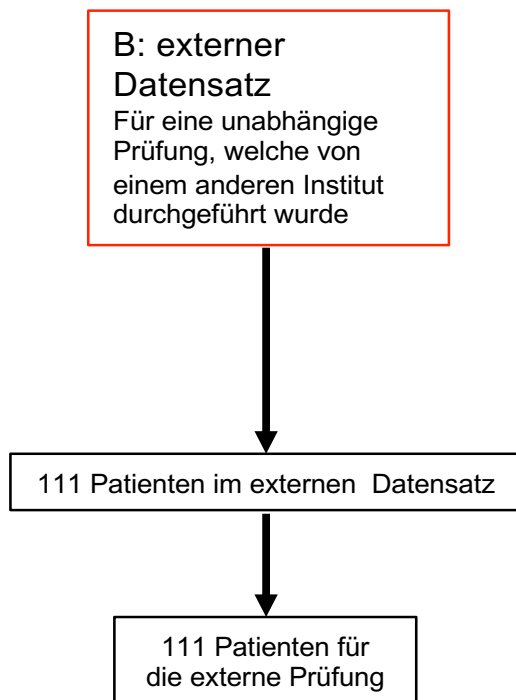
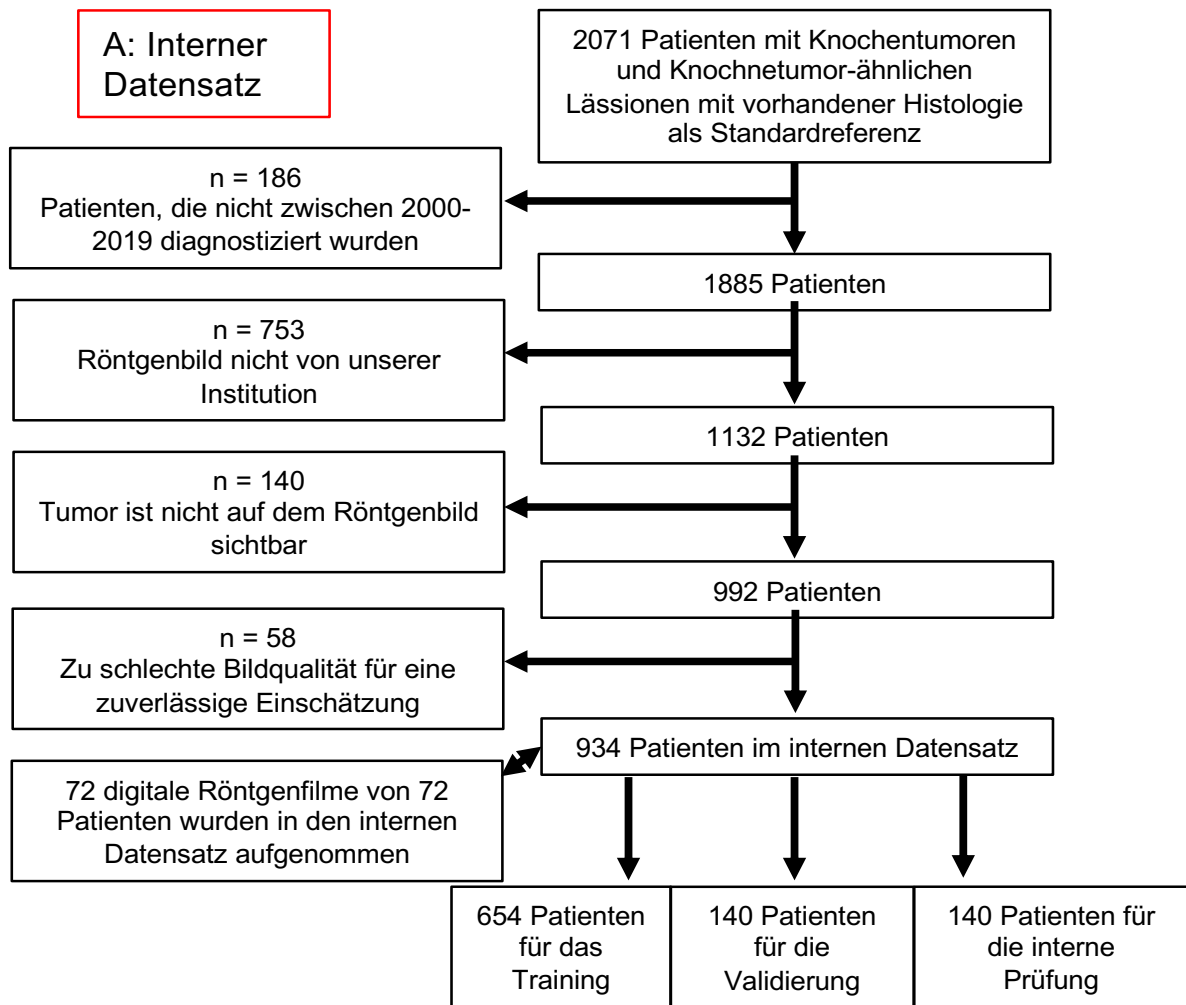


Abbildung 7: Flussdiagramm der Patientenauswahl

## 3.2 Modelltraining

Für das Modelltraining wurden Segmentierungen und Bounding-Box-Platzierungen der Tumore mit Hilfe der Open-Source-Software 3D Slicer (Version 4.7; [www.slicer.org](http://www.slicer.org)) durchgeführt. Als Segmentierung bezeichnet man die tatsächliche Kennzeichnung der Ausdehnung des Tumors auf dem Röntgenbild. Bounding-Boxen hingegen indizieren nur den Bereich auf dem Röntgenbild, der Auffälligkeiten aufweist. Dies fand ohne Kenntnis von histopathologischen und klinischen Daten statt. Um die Segmentierung durchführen zu können, wurden die 934 Röntgenbilder, welche als PNG-Datei gespeichert wurden, mit dem Programm 3D Slicer bearbeitet. Dabei wurde der Tumor in seiner ganzen Ausdehnung am äußeren Rand markiert (vgl. *Abbildung 8*). Die Speicherung der Segmentierung erfolgte als PNG Dokument und als `seg.nrrd`. Um die Zuverlässigkeit der Segmentierung zu überprüfen, wurde drei Monate nach der ersten Segmentierung eine weitere Segmentierung von 50 zufällig ausgewählten Röntgenbildern aus dem internen Datensatz durchgeführt.

Die Befundung der externen Röntgenbilder fand durch zwei Assistenzärzt\*innen für Radiologie und zwei Fachärzt\*innen für Radiologie mit muskuloskelettalem Schwerpunkt über 2 Tage hinweg statt. Die Radiologen klassifizierten jeden Tumor auf den Röntgenbildern als einen malignen oder benignen Tumor und nahmen zusätzlich eine Unterklassifizierung der vorliegenden Entität vor.

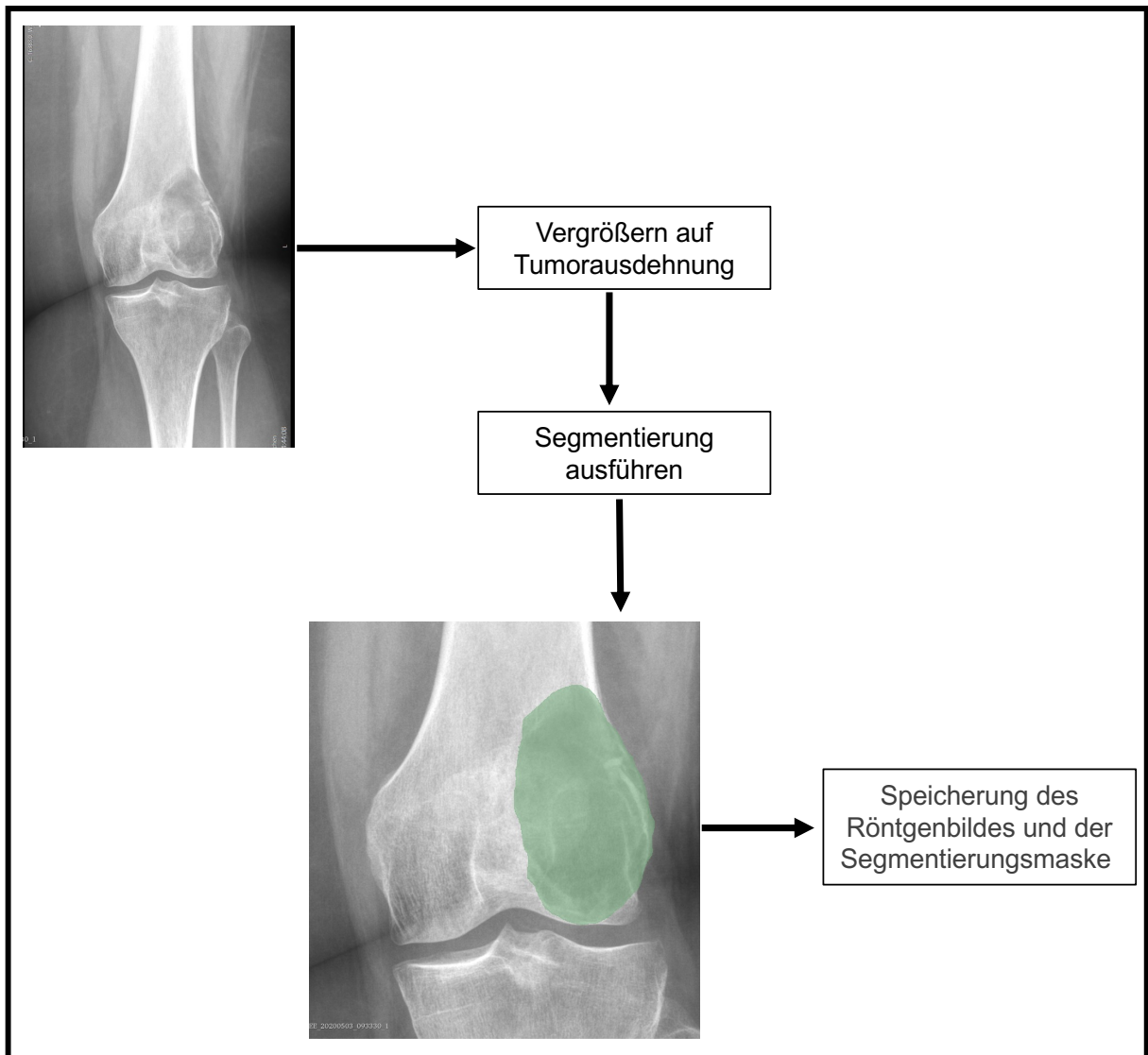


Abbildung 8: Darstellung der Segmentierung

### 3.3 Modell-Architektur

Die Durchführung der DL-Studie wurde auf einem Computer mit einem Intel Core i9-9900K mit 3.60GHz Taktrate und 16 Threads durchgeführt. Diese Zentraleinheit verfügt über 32GB DDR4 SDRAM Hauptspeicher. Hierbei steht DDR4 SDRAM für Double Data Rate 4 Synchronous Dynamic Random Access Memory und repräsentiert somit eine gängige Art von Hauptspeichern. Ferner wurde eine GeForce RTX 2080 Ti Grafikkarte von Nvidia verwendet. Das zugrundeliegende Betriebssystem war ein Linux-System (Ubuntu, Version 18.04) mit einer handelsüblichen CUDA Programmierschnittstelle (Version 11.0 von Nvidia). Eine CUDA Programmierschnittstelle ermöglicht hierbei, dass Programmteile durch den Prozessor der Grafikkarte (GPU) abgearbeitet werden können und somit hochgradig parallelisierte Programmabläufe möglich sind.

Die Modellimplementierung erfolgte in der open-source Software Python (Version 3.8.3, Python Software Foundation), wobei folgende open-source Bibliotheken verwendet wurden: PyTorch (Version 1.7) und Dectron2 (Version 0.3) (Paszke A, 2017) (Wu Y, 2019). Grundsätzlich basiert das Modell auf einem Microsoft Common Objects in Context (COCO) gefaltetem neuronalen Netzwerk (Mask-RCNN-X101) (Lin TY, 2014) (Pan SJ, 2010).

Grundsätzlich schlägt das auf dem Objektdetektor beruhende Modell viele Bounding Boxen mit zugehörigen Klassen vor, von welchen das DL-Modell die wahrscheinlichste Klasse auswählt. Dieses Vorgehen lässt sich jedoch nicht in eine binäre Klassifikationsaufgabe konvertieren, wodurch eine Operator-Empfänger-Charakteristik abgeleitet werden kann, um dann die entsprechenden optimalen Schwellenwerte zu bestimmen. Der Sigmoid Verlust pro Pixel und der binäre Verlust pro Pixel, der Regressionsverlust und der kategoriale Kreuzentropieverlust stellen die verwendeten Verlustfunktionen dar, welche für die Segmentierung, Platzierung der Bounding Boxen und Klassifikation verwendet wurden. Die Verarbeitungszeit pro Bild bei der Inferenz betrug 0.29 Sekunden. *Abbildung 9* visualisiert eine vereinfachte Darstellung der Modell-Architektur.

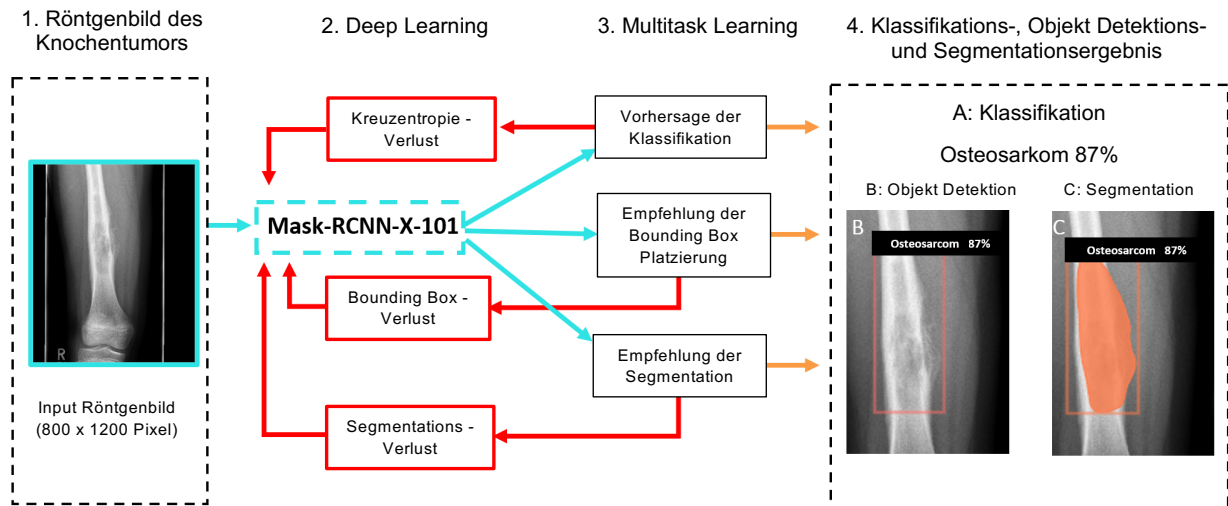


Abbildung 9: Vereinfachte Darstellung der Modell-Architektur (von Schacky CE, 2021)

### 3.4 Statistische Analyse

Um die Leistung des Multitask-Deep-Learning Modells für die Platzierung der Bounding-Boxen auf den Validierungsset und Testset zu bewerten, wurde die Intersection over Union (IoU) genutzt. IoU ist die am häufigsten verwendete Metrik für den Vergleich der Ähnlichkeit zwischen zwei beliebigen Formen und kann so für die Bewertung der Genauigkeit der Bounding Boxen eingesetzt werden (Hamid, 2019). Von einer korrekten Platzierung der Bounding Boxen wurde ausgegangen, wenn der IoU-Wert größer als 0,5 war. Eine schematische Visualisierung ist in *Abbildung 10* dargestellt. Die Leistung des DL-Modells für die Segmentierungen wurde anhand des Dice-Scores bewertet. Der Dice-Score gibt hierbei die Übereinstimmung von Größe und Lokalisation bei der Segmentierung von Objekten an (Eelbode T, 2020). Die beiden genannten Metriken IoU und Dice-Score sind miteinander vergleichbar, unterscheiden sich zwar in der Definition, können jedoch ineinander umgerechnet werden. Die endgültige Klassifizierungsleistung wurde anschließend anhand der Konfusionsmatrizen bzw. Wahrheitsmatrix, der Genauigkeit, der Sensitivität und der Spezifität mit scikit-learn, Version 0.22.2 (<https://scikit-learn.org/stable/>), wie auch von Pedregosa et al. definiert (Pedregosa, et al., 2011), für den internen und externen Testsatz bewertet. Für den Leistungsvergleich wurde der McNemar-Test verwendet, indem ein P-Wert von weniger als 0,05 als einen signifikanten Unterschied gewertet wurde. Der McNemar-Test gehört zur Klasse der Chi-Quadrat-Tests und vergleicht zwei Stichproben in Bezug auf ein dichotomes Merkmal (Twisk, 2003). Mit der Open-Source-Software Statsmodels (Version 0.9; <https://www.statsmodels.org/stable/index.html>) wurden sowohl der linear gewichtete Cohen k-Wert als auch die 95% CL berechnet (McHugh ML, , 2012). Der Cohen k-Wert ist ein Maß für die Übereinstimmung zweier verbundener kategorialer Stichproben (Twisk, 2003).

Das Modelltraining sowie die Modellbewertung wurden von zwei Autoren durchgeführt (C.E.v.S., 8 Jahre Erfahrung in der Datenanalyse; N.J.W., Statistiker und Informatiker mit 8 Jahren Erfahrung in Statistik und Datenanalyse).

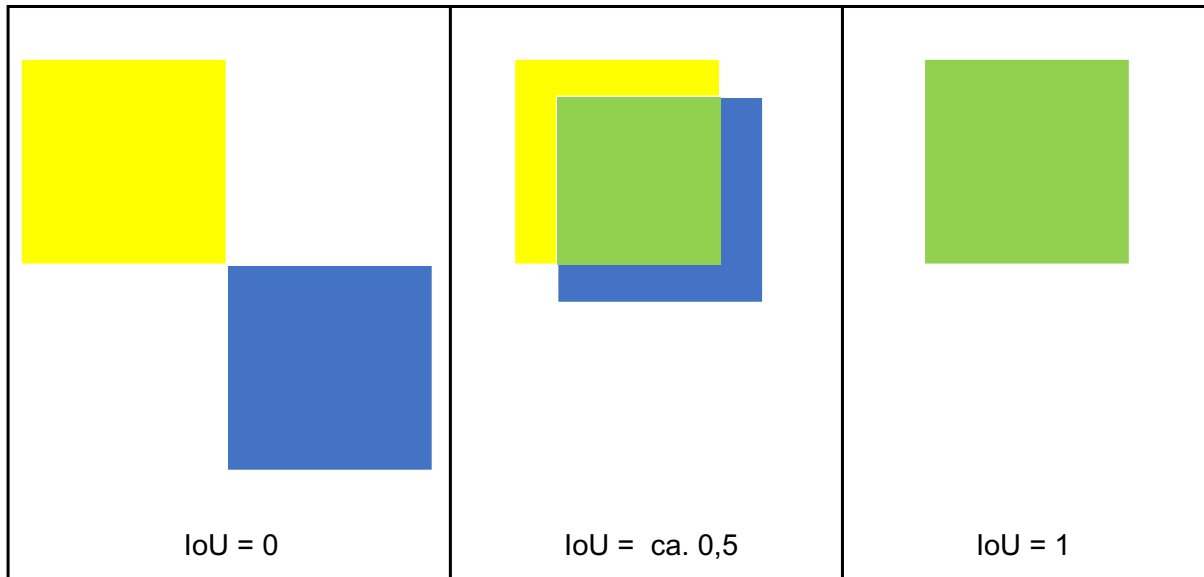


Abbildung 10: Schematische Visualisierung der Intersection over Union Kennzahl

## 4 Ergebnisse

### 4.1 Patientenmerkmale und Datensätze

Wie in Tabelle 5 und *Tabelle 6* dargestellt wurden 934 Röntgenbilder aus dem internen Datensatz verwendet. Das Durchschnittsalter betrug 33 Jahre  $\pm$  19 Jahre (Standardabweichung). Davon waren 419 Patientinnen weiblich. Aus den histologischen Befunden geht hervor, dass 667 Röntgenbilder mit benignen Knochentumoren und 267 Röntgenbilder mit malignen Knochentumoren mit einbezogen wurden. Der externe Datensatz umfasste 111 Röntgenbilder von 111 Patient\*innen mit einem Durchschnittsalter von 31 Jahre  $\pm$  22 Jahre (Standardabweichung). Davon waren 46 Patientinnen weiblich. Weitere Einzelheiten, wie zum Beispiel das prozentuale Vorkommen der einzelnen Tumorentitäten, sind in Tabelle 5 zu finden. Zu den Chondrosarkomen zählen die atypischen Knorpeltumore Grad I (n = 10), Chondrosarkome des Grades II (n = 57) und Chondrosarkome des Grades III (n = 35).

Die Intrarater-Zuverlässigkeit bei der Platzierung der Bounding Boxen betrug  $0,88 \pm 0,13$ . Dies wurde mit der Hilfe des IoU gemessen. Für Segmentierungen betrug die Intrarater-Zuverlässigkeit  $0,93 \pm 0,12$ , dies wurde anhand des Dice-Scores gemessen.

<b>Merkmale</b>	<b>Gesamt (n=934)</b>	<b>Trainingsset (654/934, 70%)</b>	<b>Validationsset (140/934, 15%)</b>	<b>Testset (140/934, 15%)</b>	<b>Externes Testset (n= 111)</b>
Alter in Jahren	33 ± 19	34 ± 20	32 ± 17	30 ± 18	33 ± 22
Geschlecht (weiblich)	419 (50%)	309 (47%)	63 (45%)	47 (34%)	46 (41%)
<b>Maligne Untergruppe</b>	267 (29%)	189 (29%)	39 (28%)	39 (28%)	35 (32%)
Chondrosarkom	102 (11%)	72 (11%)	15 (11%)	15 (11%)	11 (10%)
Osteosarkom	48 (5%)	34 (5%)	7 (5%)	7 (5%)	7 (6%)
Ewing Sarkom	45 (5%)	31 (5%)	7 (5%)	7 (5%)	6 (5%)
Plasmazell Myelom	30 (3%)	22 (3%)	4 (3%)	4 (3%)	6 (5%)
B-Zell NHL	36 (4%)	26 (4%)	5 (4%)	5 (4%)	5 (5%)
Chordom	6 (1%)	4 (1%)	1 (1%)	1 (1%)	0 (0%)
<b>Benigne Untergruppe</b>	667 (71%)	465 (71%)	101 (72%)	101 (72%)	76 (68%)
Osteochondrom	228 (24%)	160 (24%)	34 (24%)	34 (24%)	17 (15%)
Enchondrom	153 (16%)	107 (16%)	23 (16%)	23 (16%)	13 (12%)
Chondroblastom	19 (2%)	13 (2%)	3 (2%)	3 (2%)	3 (3%)
Riesenzelltumor	44 (5%)	30 (5%)	7 (5%)	7 (5%)	7 (6%)
Osteoid Osteom	19 (2%)	13 (2%)	3 (2%)	3 (2%)	4 (4%)
Nicht-ossifizierendes Fibrom	34 (4%)	24 (4%)	5 (4%)	5 (4%)	5 (4%)
Hämangiom	12 (1%)	8 (1%)	2 (1%)	2 (1%)	3 (3%)
Aneurysmatische Knochenzyste	82 (9%)	58 (9%)	12 (9%)	12 (9%)	9 (8%)
Knochenzyste	24 (3%)	16 (2%)	4 (3%)	4 (3%)	7 (6%)
Fibröse Dysplasie	52 (6%)	36 (6%)	8 (6%)	8 (6%)	6 (5%)

Tabelle 5: Verteilung der Patient\*innen



Merkmale	Gesamt (n=934)	Trainingsset (654/934, 70%)	Validationsset (140/934, 15%)	Testset (140/934, 15%)	Externes Testset (n= 111)
<b>Lokalisation gesamt</b>					
Kopf/Rumpf	125 (13%)	90 (14%)	17 (12%)	18 (13%)	6 (5%)
Extremitäten	809 (87%)	564 (86%)	123 (88%)	122 (87%)	105 (95%)
<b>Lokalisation maligne Tumore</b>					
Kopf/Rumpf	42 (16%)	28 (15%)	7 (18%)	7 (18%)	2 (6%)
Extremitäten	225 (84%)	161 (85%)	32 (82%)	32 (82%)	33 (94%)
<b>Lokalisation benigne Tumore</b>					
Kopf/Rumpf	83 (12%)	62 (13%)	10 (10%)	11 (11%)	4 (5%)
Extremitäten	584 (88%)	403 (87%)	91 (90%)	90 (89%)	72 (95%)

Tabelle 6: Fortsetzung der Tabelle 5: Verteilung der Patient\*innen

## 4.2 Modellleistung bei der internen Testreihe

In der finalen internen Testreihe nach Training und Validierung erreichte das Multitask-DL-Modell eine Genauigkeit von 81,4 % (114 von 140; 95 % Konfidenzintervall (CL): 75,0 %, 87,8 %), eine Sensitivität von 59,0 % (23 von 39; 95 % CL: 44 %, 77 %) und eine Spezifität von 90,1 % (91 von 101; 95 % CL: 84,3 %, 95,9 %) für die Klassifizierung von Knochentumoren als benigne oder maligne. Diese Ergebnisse sind in *Abbildung 11* als Konfusionsmatrix für die Unterscheidung von benignen und malignen Tumoren visualisiert worden. Die Sensitivität bezeichnet den Anteil der wirklichen malignen Tumoren, die korrekt als maligne vorhergesagt werden. Die Spezifität ist somit der Anteil der wirklichen benigne Tumoren, die korrekt als benigne vorhergesagt werden (Powers, DMW, 2011). Die Übereinstimmung zwischen der Bewertung des Multitask-DL-Modells und der histologischen Ergebnisse für die Klassifizierung der Knochentumore als benigne oder maligne betrug gemessen anhand des linear gewichtet Cohen k-Wertes 0,52 (95 % CL: 0,44, 0,60). Das Modell erreichte eine Genauigkeit von 42,9 % (60 von 140; 95 % CL: 34,7 %, 51,1 %) für die Subklassifikation der Tumorentitäten in eine von 16 Tumorentitäten und eine Makro-F1-Score von 0,33.

Der Makro-F1-Score wird verwendet, um mehrere Klassen, die zur Auswahl stehen, zu bewerten (Farkaš I, 2021). Dabei ist 1.0 der beste Wert und 0.0 der schlechteste.

Das Multitask-DL-Modell platzierte bei der Objekterkennung 65,0 % (91 von 140) der Bounding Boxes korrekt ( $> 0,5$ ). Dies ergab einen IoU von  $0,54 \pm 0,32$ . Die Segmentierungen, welche vom Multitask-DL-Modell durchgeführt wurden, ergaben einen durchschnittlichen Dice-Score von  $0,63 \pm 0,34$  mit dem internen Testset.

Histologische Standard Referenz	benigne	91	10
	maligne	16	23
		benigne	maligne

Bewertung des Deep-Learning Modells

Abbildung 11: Konfusionsmatrix der internen Testreihe. Die Zahlen entsprechen den verschiedenen Vorhersagen (von Schacky CE, 2021).

#### 4.3 Leistung des Modells in der externen Testreihe

Das Multitask-DL-Modell erreichte in der externen Testreihe eine Genauigkeit von 80,2% (89 von 111; 95% CL: 72,8%, 87,6%), eine Sensitivität von 62,9% (22 von 35; 95% CL: 47%, 79%) und 88,2% Spezifität (67 von 76; 95% CL: 81%, 96%) für die Klassifizierung von Knochentumoren als benigne oder maligne. Diese Ergebnisse sind in *Abbildung 12* als Konfusionsmatrix für die Unterscheidung von benignen und malignen Tumoren dargestellt worden.

Die Übereinstimmung zwischen der Bewertung des Multitask-DL-Modells und der histologischen Ergebnisse für die Klassifizierung der Knochentumore als benigne oder maligne betrug gemessen anhand des linear gewichtet Cohen k-Werte 0,53 (95 % CL: 0,43, 0,62). Das Modell erreichte eine Genauigkeit von 43,2 % (48 von 111; 95 % CL: 34,0 %, 52,4 %) für die Subklassifikation der Tumorentitäten in eine von 16 Tumorentitäten und eine Makro-F1-Score von 0,34.

Das Multitask-DL-Modell platzierte bei der Objekterkennung 59,5 % (66 von 111) der Bounding Boxes korrekt ( $\text{IoU} > 0,5$ ). Dies ergab einen IoU von  $0,52 \pm 0,34$ . Eine zumindest geringe Überlappung ( $\text{IoU} > 0$ ) zwischen Bounding-Box und den Tumoren

gab es in 82,0% (91 von 111). Die Segmentierungen, welche vom Multitask-DL-Modell durchgeführt wurden, ergaben einen durchschnittlichen Dice-Score von  $0,60 \pm 0,37$  mit dem externen Testset. *Tabelle 7* gibt einen Überblick über die Leistung des DL-Modells bei der Identifizierung maligner Läsionen in den internen und externen Testgruppen sowie bei der Klassifizierung von Tumorentitäten, der Platzierung von Bounding Boxen und der Segmentierung.

Wenn eine gewisse Überlappung ( $IoU > 0$ ) festgestellt wurde, klassifizierte das Modell 85,7% (78 von 91) als korrekt benigne oder maligne. Wurde keine Überlappung festgestellt ( $IoU = 0$ ), klassifizierte das Modell 55,0% (11 von 20) als korrekt benigne oder maligne.

Hinsichtlich der Lokalisation erreichte das Modell 50,0 % Genauigkeit (drei von sechs) bei Tumoren in Rumpf- und Kopfbereich, 78,8 % Genauigkeit (26 von 33) im Bereich der oberen Extremitäten und 83,3 % (60 von 72) im Bereich der unteren Extremitäten.

Der Zusammenhang zwischen Tumorgöße und Genauigkeit wurde untersucht, indem die Tumore je nach Größe in zwei gleichgroße Gruppen eingeteilt wurden. Die Genauigkeit bei den kleinen Tumoren ( $<$  als 6,4 cm im maximalen Durchmesser auf dem Röntgenbild) betrug 78,2% (43 von 55), hingegen betrug die Genauigkeit bei großen Tumoren ( $>$  als 6,4 cm im maximalen Durchmesser) 82,1% (46 von 56).

Histologische Standard Referenz	benigne	67	9
	maligne	13	22
		benigne	maligne

Bewertung des Deep-Learning Modells

Abbildung 12: Konfusionsmatrix der externen Testreihe. Die Zahlen entsprechen den verschiedenen Vorhersagen (von Schacky CE, 2021).

Multitask-DL-Modell	Internes Testset	Externes Testset
<b>Genauigkeit der Malignitätsprognose</b>	81.4 (114/140) [75.0, 87.8]	80.2 (89/111) [72.8, 87.6]
<b>Sensitivität der Malignitätsprognose</b>	59.0 (23/39) [44, 77]	62.9 (22/35) [47, 79]
<b>Spezifität der Malignitätsprognose</b>	90.1 (91/101) [84.3, 95.9]	88.2 (67/76) [81, 96]
<b>Genauigkeit der 16 Tumorentität</b>	42.9 (60/140) [34.7, 51.1]	43.2 (48/111) [34.0, 52.4]
<b>Intersection over Union</b>	0.54 ± 0.32	0.52 ± 0.34
<b>Dice Score</b>	0.63 ± 0.34	0.60 ± 0.37

Hinweis: Die Daten für Genauigkeit, Sensitivität und Spezifität sind als Prozentsätze ausgedrückt, wobei die Proportionen in den runden Klammern angegeben sind. Die in den eckigen Klammern aufgeführten Daten sind 95% CLs. Die Daten für die IoU und den Dice-Score werden als Mittelwert ± Standardabweichung angegeben. Sensitivität und Spezifität wurden ohne Berücksichtigung von Patient\*innen mit normalem histologischem Befund berechnet.

Tabelle 7: Leistung des DL-Modells bei der Identifizierung maligner Tumore in den internen und externen Testgruppen sowie bei der Klassifizierung von Tumorentitäten, der Platzierung von Bounding Boxen und der Segmentierung (von Schacky CE, 2021).

#### 4.4 Leistung des Modells im Vergleich mit Radiolog\*innen

Im Vergleich zum Multitask-Deep-Learning Modell erreichten zwei Assistenzärzt\*innen für Radiologie eine Genauigkeit von 71,2 % (79 von 111; 95 % CL: 57,0 %, 74,6 %) und 64,9 % (72 von 111; 95 % CI: 56,1 %, 73,9 %), eine Sensitivität von 60,0 % (21 von 35; 95 % CL: 44 %, 76 %) bzw. eine Sensitivität von 37,1 % (13 von 35; 95 % CL: 21%, 53%), außerdem eine Spezifität von 76,3 % (58 von 76; 95 % CL: 63 %, 80 %) bzw. eine Spezifität von 77,6 % (59 von 76; 95 % CL: 56 %, 74 %) für die Klassifizierung von Knochentumoren als benigne oder maligne.

Die Übereinstimmungen zwischen den Assistenzärzt\*innen für Radiologie und der histologischen Ergebnisse betragen gemessen anhand des linear gewichtet Cohen k-Wertes 0,35 (95 % CL: 0,17, 0,53) und 0,15 (95% CL: 0,0, 0,35).

Die beiden Assistenzärzt\*innen für Radiologie erreichten eine Genauigkeit von 44,1 % (49 von 111; 95 % CL: 34,9 %, 53,3 %) für die Subklassifikation der Tumorentitäten in eine von 16 Tumorentitäten.

Bei der Klassifizierung eines Tumors als benigner oder maligner war die Genauigkeit des DL-Modells höher als die der beiden Assistenzärzt\*innen für Radiologie (80,2 % gegenüber 71,2% bzw. 64,9%; P = .002 bzw. P < .001).

Im Gegensatz dazu erreichten zwei Fachärzt\*innen für Radiologie mit muskuloskelettalem-Schwerpunkt im Vergleich zum Multitask-Deep-Learning Modell

eine Genauigkeit von 83,3% (93 von 111; 95 % CL: 76,9 %, 90,7 %) bzw. 82,9 % (92 von 111; 95 % CL: 76,0 %, 90,0 %), eine Sensitivität von 88,6 % (31 von 35; 95 % CL: 79 %, 99 %) bzw. eine Sensitivität von 82,9 % (29 von 35; 95 % CL: 71 %, 95 %), außerdem eine Spezifität von 81,6 % (62 von 76; 95 % CL: 73 %, 91 %) bzw. eine Spezifität von 82,9 % (63 von 76; 95 % CL: 76 %, 90 %) für die Klassifizierung von Knochentumoren als benigne oder maligne.

Die Übereinstimmungen zwischen den Fachärzt\*innen für Radiologie mit muskuloskelettalem Schwerpunkt und der histologischen Ergebnisse betragen gemessen anhand des linear gewichtet Cohen k-Wertes 0,65 (95 % CL: 0,51, 0,80) bzw. 0,62 (95% CL: 0,47, 0,78).

Außerdem erreichten die beiden Fachärzt\*innen für Radiologie mit muskuloskelettalem Schwerpunkt eine Genauigkeit von 58,6 % (65 von 111; 95 % CL: 49,4 %, 67,8 %) bzw. 65,8 % (73 von 111; 95 % CL: 56,5 %, 73,9 %) für die Subklassifikation der Tumorentitäten in eine von 16 Tumorentitäten.

Zwischen dem DL-Modell und den beiden Fachärzt\*innen für Radiologie mit muskuloskelettalem Schwerpunkt konnte kein signifikanter Unterschied (80,2 % vs. 83,8 % und 82,9 %;  $P = 0,13$  bzw.  $P = 0,25$ ) festgestellt werden, was die genaue Klassifizierung eines Tumors in einen benignen oder malignen betrifft.

*Tabelle 8* gibt einen Überblick über den Leistungsvergleich zwischen dem DL-Modell, den Assistenzärzt\*innen für Radiologie und den Fachärzt\*innen für Radiologie mit muskuloskelettalem Schwerpunkt bei der Identifizierung maligner Tumoren und der Klassifizierung von Tumorentitäten aus der externen Testgruppe.

<b>Externes Testset</b>	<b>Multitask-DL-Model</b>	<b>Assistenzärzt*innen für Radiologie</b>	<b>Fachärzt*innen für Radiologie mit muskuloskeletalem Schwerpunkt</b>
<b>Genauigkeit der Malignitätsprognose</b>	80.2 (89/111) [72.9, 87.6]	68.0 ± 3.2 [61.6, 73.8]	83.4 ± 0.5 [77.9, 87.8]
<b>Sensitivität der Malignitätsprognose</b>	62.9 (22/35) [47, 79]	49 ± 11 [37, 60]	86 ± 2.9 [76,92]
<b>Spezifität der Malignitätsprognose</b>	88.2 (67/76) [81, 96]	77 ± 0.7 [70, 83]	82 ± 0.7 [75, 88]
<b>Genauigkeit der 16 Tumorentität</b>	43.2 (48/111) [34, 52]	44.1 ± 0.0 [34.9, 53.3]	62.2 ± 3.6 [55.6, 68.3]

Hinweis: Die Daten für Genauigkeit, Sensitivität und Spezifität werden als Prozentsätze ausgedrückt, wobei die Proportionen in den runden Klammern angegeben sind oder als mittlere Prozentsätze mit 6 Standardabweichungen. Die in den eckigen Klammern angegebenen Daten sind 95% CLs. Sensitivität und Spezifität wurden ohne Berücksichtigung von Patient\*innen mit normalem histologischem Befund berechnet.

Tabelle 8: Leistung des DL-Modells, von Assistenzärzt\*innen für Radiologie und von Fachärzt\*innen für Radiologie mit muskuloskeletalem Schwerpunkt bei der Identifizierung maligner Tumoren und der Klassifizierung von Tumorentitäten in der externen Testgruppe (von Schacky CE, 2021).

#### 4.5 Modellinterpretation und Beispiele

Abbildung 13 zeigt ein Beispiel für die korrekte Beurteilung der Tumorentität bei einem 36-jährigen Mann mit einem Chondrosarkom des linken Femurs. (A) zeigt ein natives Röntgenbild im anterior-posterior Strahlengang. (B) zeigt die korrekte Bounding-Box-Platzierung und Segmentierung des Tumors durch das Multitask-Deep-Learning-Modell. Mit einer Sicherheit von 97% klassifiziert das DL-Modell diesen Tumor korrekt als Chondrosarkom.

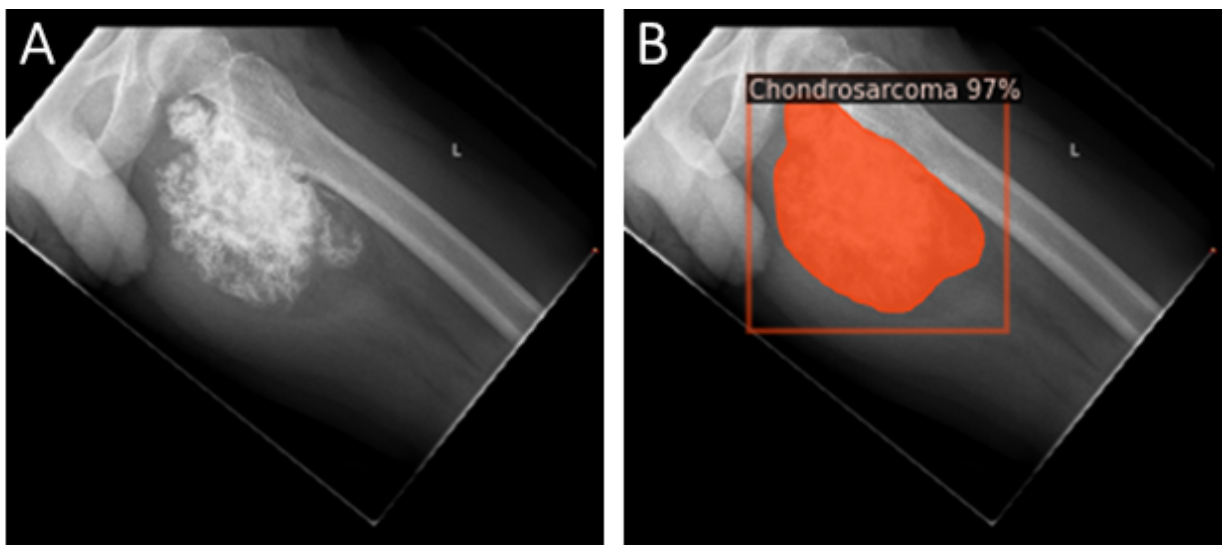


Abbildung 13: Beispiel für die korrekte Bewertung als Chondrosarkom (von Schacky CE, 2021)

Abbildung 14 zeigt ein weiteres Beispiel für die korrekte Beurteilung der Tumorentität bei einem 16-jährigen Jungen mit einem nichtossifizierenden Fibrom (NOF) der linken Fibula. (A) zeigt ein natives Röntgenbild im anterior-posterior Strahlengang. (B) zeigt die korrekte Bounding-Box-Platzierung und Segmentierung des Tumors durch das Multitask-Deep-Learning-Modell. Mit einer Sicherheit von 72% klassifiziert das DL-Modell diesen Tumor korrekt als NOF.

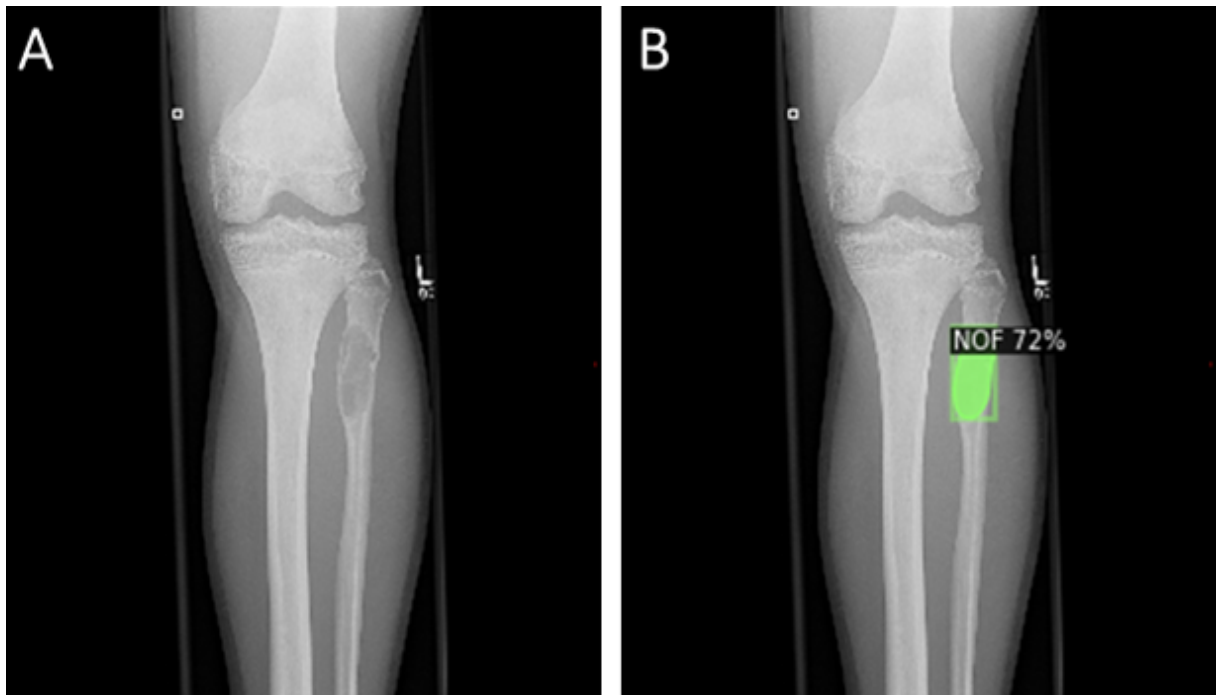


Abbildung 14: Beispiel für die korrekte Bewertung als NOF (von Schacky CE, 2021)



Abbildung 15 zeigt ein Beispiel für die korrekte Bewertung des Tumors als benigne aus dem externen Testset bei einem 39-jährigen Mann mit einem Enchondrom des linken Femurs. (A) zeigt ein natives Röntgenbild im anterior-posterior Strahlengang. (B) zeigt die korrekte Bounding-Box-Platzierung und Segmentierung (grün) des Tumors (blau) durch das Multitask-Deep-Learning-Modell. Mit einer Sicherheit von 89 % klassifiziert das DL-Modell diesen Tumor korrekt als benigne.

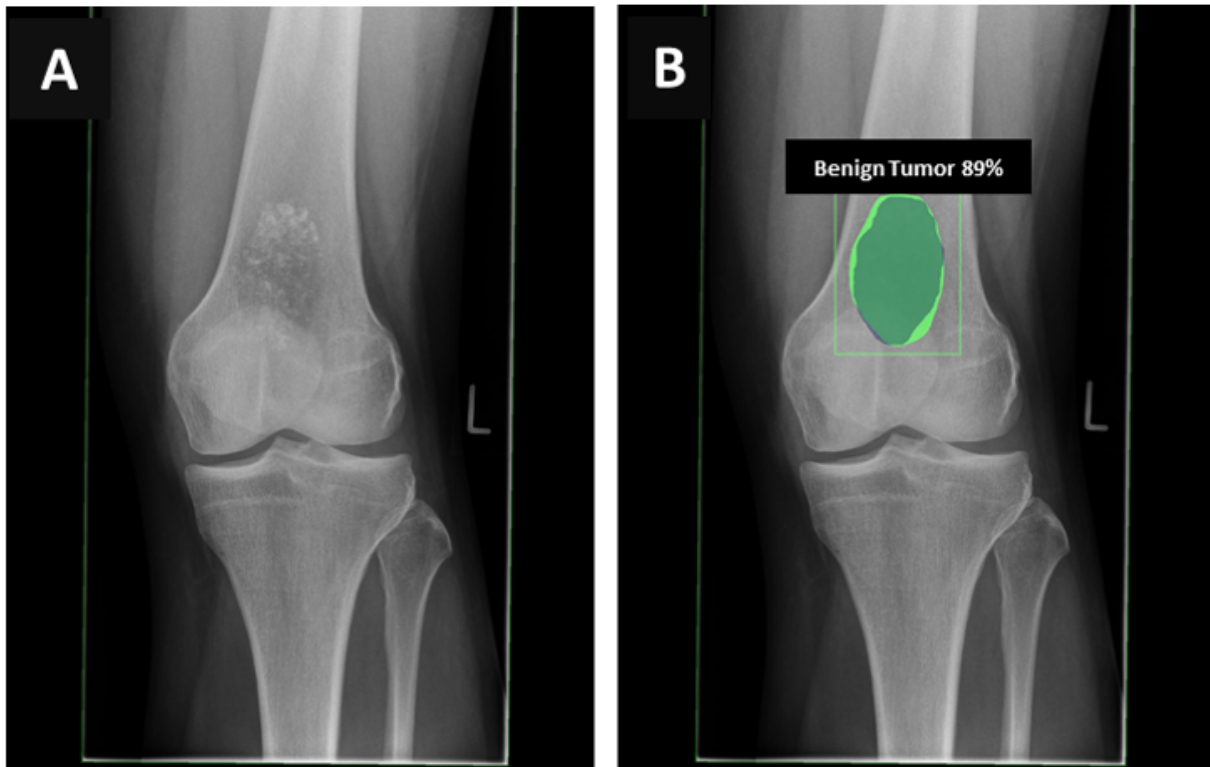


Abbildung 15: Beispiel für die korrekte Klassifikation als benigner Tumor (von Schacky CE, 2021)

Abbildung 16 zeigt ein Beispiel für die korrekte Bewertung des Tumors als maligne bei einem 87-jährigen Mann mit B-Zell-Non-Hodgkin-Lymphom. (A) zeigt ein natives Röntgenbild im anterior-posterior Strahlengang. (B) zeigt die korrekte Bounding-Box-Platzierung und Segmentierung (rot) der Knochenläsion (orange) durch das Multitask-Deep-Learning-Modell. Mit einer Sicherheit von 86 % klassifiziert das DL-Modell diesen Tumor korrekt als malignen.

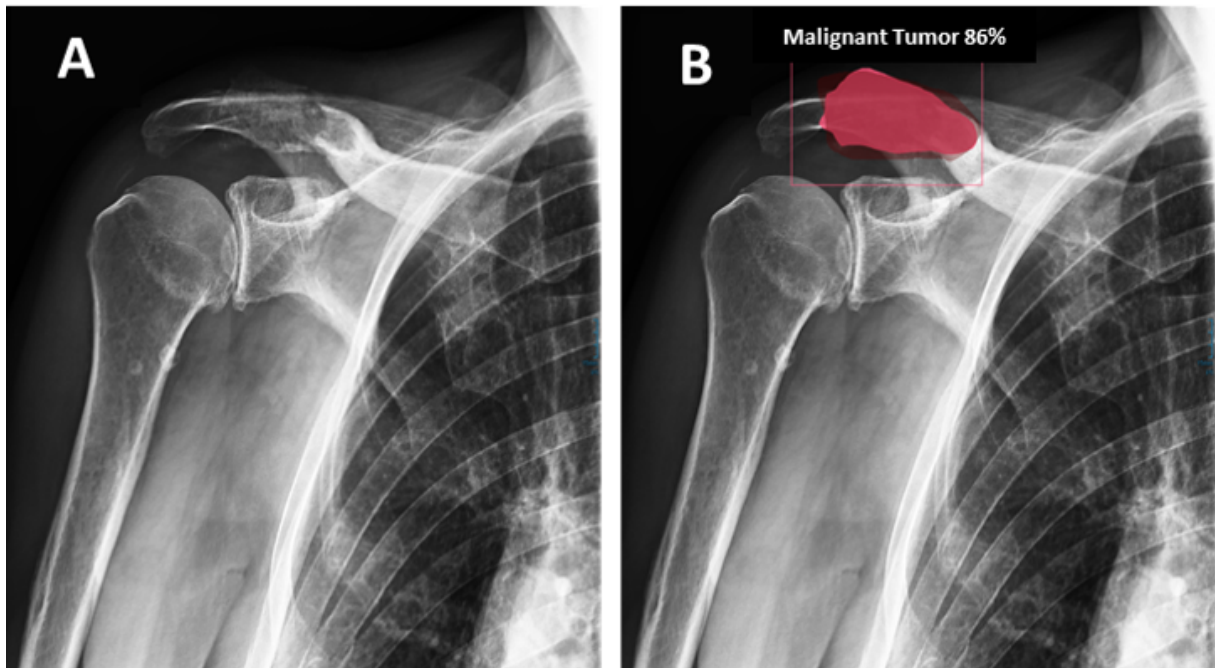


Abbildung 16: Beispiel für die korrekte Klassifikation als maligner Tumor (von Schacky CE, 2021)

Abbildung 17 zeigt ein Beispiel für eine falsche Vorhersage durch das Multitask-Deep-Learning-Modell aus dem externen Testset bei einem 17-jährigen Jungen mit einer aneurysmatischen Knochenzyste im linken Femur. (A) zeigt ein natives Röntgenbild im anterior-posterior Strahlengang. (B) zeigt die Vorhersage der Bounding-Box-Platzierung und der Segmentierung (orange) des DL- Modells, welches den Tumor mit einer Sicherheit von 63% als maligne einstuft. Diese fehlerhafte Vorhersage kann durch Überlagerung von Darmschlingen verursacht worden sein. Es handelt sich eigentlich um einen benignen Tumor im linken Femur (blau).

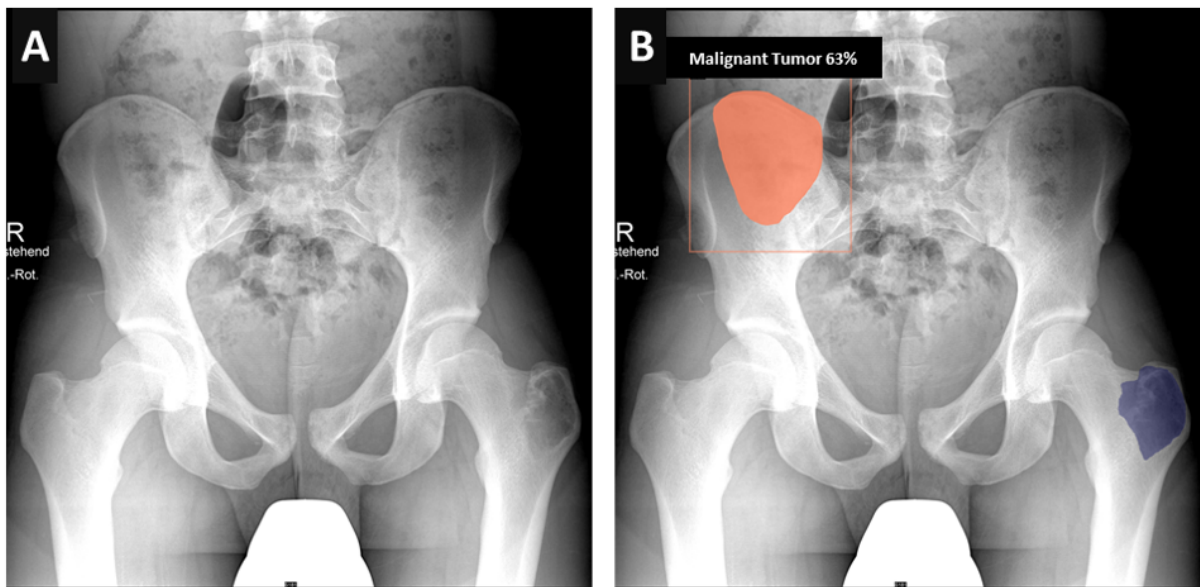


Abbildung 17: Beispiel für eine fehlerhafte Bewertung (orange Segmentierungs-Maske). Die eigentliche Läsion befindet sich im Trochanter major links/proximalen Femur links (lila Segmentierungs-Maske) (von Schacky CE, 2021)

## 5 Diskussion

Um eine Verbesserung der Erkennungsraten von primären Knochentumoren zu erlangen, ist ein Multitask-Deep-Learning (DL)-Modell, das Bounding-Box-Platzierungen, Segmentierungen und Klassifizierungen von Knochentumoren auf Röntgenbildern durchführt, sehr hilfreich. Daher wurde ein Multitask-DL-Modell entwickelt, welches Bounding-Box-Platzierungen, Segmentationen und Klassifizierung von Knochentumoren auf Röntgenbildern vornimmt. Das entwickelte DL-Modell platzierte die Bounding-Boxen der Tumore mit einer beinahe vollständigen Überlappung. Mit einem entsprechend ermittelten Dice-Wert führte das DL-Modell Segmentierungen durch, welche eine große Übereinstimmung mit denjenigen lieferten, die von Radiolog\*innen durchgeführt wurden und als ‚ground truth‘ dienten. Bei der Klassifizierung von Knochentumoren als benigne oder maligne übertraf die Genauigkeit des Modells die der Assistenzärzt\*innen für Radiologie. Die Genauigkeit des DL-Modells war hingegen mit Fachärzt\*innen für Radiologie mit muskuloskelettalem-Schwerpunkt vergleichbar. Dieses Ergebnis zeigt aussichtsreiche Perspektiven für einen Einsatz des Deep-Learning Modells in der zukünftigen klinischen Anwendung.

Zur Beurteilung von Knochentumoren wurden in früheren Studien Informationen aus Röntgenbildern verwendet. Lodwick et al. erstellten ein bekanntes Klassifizierungssystem zur Einteilung von Knochenläsionen in Abhängigkeit von ihren Destruktionsmustern (Lodwick, 1964), welches auch heute noch in der Radiologie seine Anwendung findet. Im Jahr 2016 wurde von Caracciolo et al. eine modifizierte Lodwick-Madewell-Klassifikation vorgeschlagen. Diese soll helfen, das Malignitätsrisiko von Knochenläsionen auf Röntgenbildern besser einzuteilen (Caracciolo JT, 2016). Es wurde festgestellt, dass 94 % der Läsionen des Grades I benigne waren und 81 % der Läsionen des Grades III maligne waren. Hingegen konnte den Läsionen Grad II keine eindeutige Differenzierung zugeordnet werden, die Läsionen waren zu gleichen Teilen benigne bzw. maligne. In einer Studie von Reinus et al. wurden ein neuronales Netzwerk genutzt, um fokale Läsionen auf Röntgenbilder zu beurteilen. Es wurden 709 Knochentumore untersucht und zahlreiche demografische und röntgenologische Daten in das neuronale Netz eingegeben, was zu einer Genauigkeit von 82 % bei der Unterscheidung von benignen und malignen Läsionen führte (Reinus WR, 1994).

In einer weiteren Studie wurden 18 Merkmale, darunter Alter, Geschlecht und röntgenologische Daten, als Input für ein naives Bayes'sches Modell zur Bewertung von 710 Knochentumoren verwendet. Diese Studie erreichte bei der Bewertung von 710 Knochentumoren eine Genauigkeit von 44 % bei der Klassifizierung der Tumorentität (Do BH, 2017).

Bluemke et al. empfehlen eine Validierung, einen Hold-out-Test und eine externe Validierung durchzuführen (Bluemke DA, 2020). Diese wurden in den beiden oben genannten Studien jedoch nicht durchgeführt. Darüber hinaus wurden in diesen

Studien radiologische Merkmale als Input für die Modelle verwendet, so dass Readings und Auswertungen von spezialisierten Radiolog\*innen vorab erforderlich waren.

Der Vorteil eines vollautomatischen Designs, das sich auf ein Multitask-DL-Modell stützt, ist, dass keine Readings oder Annotationen von Radiolog\*innen benötigt werden. Das Multitask-DL-Modell, welches von uns entwickelt wurde, führt nicht nur eine automatische Objekterkennung und Tumorsegmentierung durch und zeigt so die Lage des Tumors mit nur dem Röntgenbild als Input an, sondern erreicht auch eine höhere Genauigkeit als frühere Studien. Ein solches Modell kann die Wahrscheinlichkeit einer Fehlklassifikation eines primären Knochentumors verringern und zusätzliche wichtige Informationen liefern, welche eine Überweisung an ein spezialisiertes Zentrum für Knochentumore für weitere diagnostische Untersuchungen wie Magnetresonanztomographie (MRT) oder Computertomographie (CT) erforderlich machen würde. Dies ist vor allem im klinischen Umfeld wie z.B. in kleinen Krankenhäusern und Praxen, in denen die Expertise in der muskuloskelettalen Bildgebung begrenzt ist, besonders wichtig. Die Verwendung von KI im klinischen Alltag soll Radiolog\*innen vor allem bei der Befundung von seltenen Erkrankungen wie Knochentumoren, bei denen keine ausreichende Expertise erworben werden kann, unterstützen. Lakhani et al. schreiben in ihrer Studie, dass KI einige Aufgaben der Radiolog\*innen ersetzen werde, aber in vielen Fällen wird KI die Radiolog\*innen in der genaueren Befundung vielmehr unterstützen (Lakhani P, 2017).

Die Beurteilung von Knochentumoren mittels Röntgenaufnahmen wurde durch Lodwick et al. wesentlich verbessert. Er führte einen Algorithmus ein, der eine genaue und praktikable Einstufung der Wachstumsrate in der täglichen Praxis der muskuloskelettalen Radiologen ermöglicht. Der Algorithmus wurde von Gwilym Lodwick selbst im Laufe seiner Karriere immer wieder verfeinert. In der ersten Veröffentlichung aus dem Jahr 1964 wurde das ursprünglich vorgeschlagene Einstufungssystem in Form einer Texttabelle dargestellt. Zusätzlich wurde in dieser ersten Arbeit von 1964 das Vorhandensein von Strebepfeiler, Codman-Dreieck und Tumorknochen bewertet (Lodwick, 1964). In seiner Iteration von 1980 konzentrierte sich Lodwick schließlich auf weniger Deskriptoren. Diese beschränkt sich auf die Anwendung des Klassifizierungsalgorithmus auf Tumore mit offensichtlicher Knochendestruktion (Lodwick GS, 1980). Das endgültige Klassifizierungssystem von Lodwick aus dem Jahr 1980 ist die Klassifizierung, auf die sich die Lehrbücher der Radiologie heute beziehen.

Seit der ursprünglichen Veröffentlichung von Lodwick wurden zwei alternative angepasste Varianten der Klassifizierung nach Lodwick vorgeschlagen. Im Jahr 1981 führte Madewell den Deskriptor eines sich verändernden Randes ein. Dieser Parameter soll auf eine stärkere biologische Aktivität hinweisen, die mit einer erhöhten Wahrscheinlichkeit und einer erhöhten Malignität eines Tumors einhergeht. Madewell unterscheidet den sich verändernden Rand (der auf seriellen Röntgenbildern ausgewertet wird) von Kombinationsmustern, bei denen Merkmale verschiedener Wachstumsgrade in derselben Läsion vorhanden sind, die zu einem einzigen Zeitpunkt

ausgewertet werden (Madewell JE, 1981). Mehr als 30 Jahre später, im Jahr 2016, veröffentlichten Caracciolo und Kollegen ein „Modified Lodwick-Madewell Grading System“, das auf der Veröffentlichung von Madewell mit zusätzlichen Anpassungen basiert. Caracciolo und Kollegen verfeinerten den Parameteränderungsspielraum weiter und adaptierten das Gradingssystem in IA, IB, II, IIIA, IIIB und IIIC. Die neue Kategorie IIIA soll nun für Läsionen mit Randveränderungen, die auf mehreren Röntgenbildern zu sehen sind, eingeführt werden. Auch sollen atypische Kombinationen, die auf Malignität hindeuten und wenn keine früheren Röntgenbilder verfügbar sind, dieser Kategorie zugeteilt werden. Diese Kategorie IIIA impliziert ein schnelles Wachstum des Knochentumors und ist damit ein Zeichen für eine mit hoher Wahrscheinlichkeit maligne Läsion. So sollte beispielsweise eine Läsion des ursprünglichen Grades IB, bei der in einem Folgeröntgenbild eine neue vollständige Kortikalispenetration beobachtet wird, in die neue Kategorie IIIA eingestuft werden. Bezüglich der Kombinationsmuster stellen Caracciolo und Kollegen fest, dass es ein Grundprinzip der Läsionsanalyse ist, dass die Ränder nach ihren aggressivsten Merkmalen klassifiziert werden.

Außerdem wird im System von Caracciolo et al. Grad IC zu Grad II, um dem mäßigen Malignitätsrisiko dieser Läsionen deutlich zu machen. Grad IIIB dieses Systems entspricht den Graden II und III nach Lodwick. Sowohl Madewell als auch Caracciolo und Kollegen haben die Definition der Wachstumsgrade IA und IB leicht vereinfacht, obwohl Caracciolo und Kollegen ausdrücklich darauf hinweisen, dass diese Grade im Wesentlichen unverändert gegenüber dem ursprünglichen Lodwick-Einstufungssystem bleiben (Caracciolo JT, 2016). Darüber hinaus ist es möglich, dass röntgenologisch verborgene Läsionen eine erhebliche Tumormasse auch ohne entsprechenden röntgenologischen Nachweis vorweisen können. Dies berücksichtigen sowohl Madewell (1981) als auch Caracciolo et al. (2016). Im System von Caracciolo und Kollegen werden diese Läsionen der Kategorie IIIC zugeordnet.

Die Zuweisung des Wachstumsgrades ist anfällig für mögliche Fehlklassifizierungen, was zu einer mäßigen Interobserver-Zuverlässigkeit führt. In einer Studie von Lodwick und Kollegen aus dem Jahr 1980 bewerteten vier Untersucher 223 Knochentumore (Lodwick GS, 1980). Dabei zeigte sich, dass die Leser bei der Zuweisung der Wachstumsgrade IA (am ehesten benigne) und III (am ehesten maligne) häufiger übereinstimmten als bei der Zuweisung von Zwischengraden IB, IC und II (Lodwick GS, 1980).

Das 1980 von Lodwick entwickelte Klassifizierungssystem ist zum Lehrbuchstandard für die Einteilung des Knochentumorwachstums geworden (Davies AM, 2009). Es wurde jedoch argumentiert, dass das ursprüngliche Lodwick-Klassifikationssystem zu komplex ist und einer Vereinfachung bedarf (Caracciolo JT, 2016). Daher haben Benndorf et al. einen Entscheidungsbaum des ursprünglichen Lodwick-Klassifizierungsalgorithmus erstellt, der in der klinischen Praxis verwendet werden kann (Benndorf M, 2021). Dieser Entscheidungsbaum ist in *Abbildung 4* dargestellt und soll bei der alltäglichen Befundung von Knochenläsionen eine Hilfestellung leisten.

Unsere Studie hat jedoch weitere Limitierungen. Erstens wurde die Klassifizierung der Knochentumore nur in benigne und maligne vorgenommen, obgleich auch einige benignen Knochentumore behandlungsbedürftig sein können und daher an ein spezialisiertes Zentrum für muskuloskelettalen Radiologie überwiesen werden sollten. Zur Erkennung dieser Knochentumore könnte die automatische Platzierung der Bounding Boxen ein wichtiges Hilfsmittel sein, um solche Tumore zu erkennen und gegebenenfalls weiter zu behandeln. Zweitens wurden in unsere Studie nur histologisch nachgewiesene Knochentumore eingeschlossen. Dadurch kann es zu einem Selektionsbias kommen. Höchstwahrscheinlich hat dieser Selektionsbias jedoch den Datensatz erschwert, da sowohl das DL-Modell als auch Radiolog\*innen typischerweise benigne Tumore leichter klassifizieren können als Tumore, welche zur Diagnose eine Histologie benötigen. Drittens wurden keine Patient\*innen mit normalen, histologisch gesicherten Befunden berücksichtigt. Ein unerfahrener Leser kann jedoch auf einen Tumor schließen, wenn eine normale Variante vorliegt. Aus diesem Grund werden normale Varianten normalerweise zur Berechnung der Sensitivität und Spezifität berücksichtigt. Viertens wird die klinische Anwendbarkeit eventuell durch die fehlenden Vorhersagen anderer Krankheiten wie Frakturen oder degenerativen Veränderungen, mit denen das Modell nicht trainiert wurde, eingeschränkt. Fünftens wurde möglicherweise die ursprüngliche Auflösung oder Genauigkeit der verwendeten Röntgenbilder, die als Input für das Modell dienten, verringert. Dies geschah aufgrund von Einschränkungen durch das Vortraining, in welchem die Pixel auf 800 x 1200 Pixel festgelegt wurden. Im Vergleich zu früheren DL-Modellen war die verwendete Eingangsauflösung von 800 x 1200 Pixeln dennoch höher (He Y, 2020) (Kroque JD, 2020). Sechstens wurden in unsere Studie nur primäre Knochentumore eingeschlossen und keine Knochenmetastasen. Nach Macedo et al. sind Knochenmetastasen jedoch viel häufiger als primäre Knochentumore (Macedo F, 2017). Daher sollten in zukünftigen größeren Studien Röntgenbilder mit Metastasen Berücksichtigung finden. Siebtens enthielt das DL-Modell keinerlei Information über das Alter der Patient\*innen oder über die Knochendichte. Achters wurde das DL-Modell nur mit Röntgenbildern trainiert, die aus einer Ebene bestand, obwohl die Radiolog\*innen zur Befundung zwei Ebenen nutzen. So könnte sich das Outcome des DL-Modells durch das Training auch in der zweiten Ebene eventuell verbessern. Des Weiteren wurden in unserer Studie nur konventionelle Röntgenbilder zur Beurteilung von Knochentumoren verwendet, obwohl die MRT-Untersuchung vor allem bei der Ausdehnung des Knochentumors eine wichtige Rolle spielt (Gemescu IN, 2019).

## 6 Zusammenfassung

Die vorliegende Arbeit basiert auf einer Studie, welche 934 befundete konventionelle Röntgenbilder mit primären Knochentumoren umfasst, die aus dem Zeitraum Januar 2003 bis Juni 2020 stammen. Diese Datenbasis diente zur Entwicklung eines Multitask-Deep-Learning-Modells. Ziel dieses ist die Platzierung von Bounding-Boxen um primäre Knochentumore, eine entsprechende Segmentierung der Läsionen und eine anschließende Klassifizierung in benigne oder maligne Läsionen.

Der zugrundeliegende Datensatz, der konventionellen Röntgenbildern von benignen und malignen Knochentumoren enthält, wurde in Trainings-, Validierungs- und Testdatensätze aufgeteilt. Darüber hinaus existierte ein weiterer externer Datensatz von einer anderen Universitätsklinik zur geografischen Validierung des entwickelten Modells. Diese Datengrundlage wurde verwendet, um ein entsprechendes Deep-Learning Modell (Convolutional Neural Network), das eine Subgruppe der Künstlichen Intelligenz darstellt, zu trainieren, zu validieren und zu testen. Die Bewertung der Modelleleistung erfolgte auf Basis einer statistischen Analyse, die u.a. die Genauigkeit, Sensitivität und Spezifität berücksichtigte. Dabei erzielte das Modell eine Spezifität von bis zu 90,1%. Die Leistung des Multitask-Deep-Learning-Modells in der Klassifizierung der Knochentumore als benigne oder maligne Läsion übertraf die Leistung der Assistenzärzt\*innen für Radiologie und erreichte eine vergleichbar gute Leistung wie die der Fachärzt\*innen für Radiologie mit muskuloskelettaler Expertise. Dieses Modell kann durch die diagnostische Genauigkeit die Befundung von primären Knochentumoren auf konventionellen Röntgenbildern verbessern und somit eine schnelle, angepasste Weiterbehandlung der Patient\*innen gewährleisten. Dies ist vor allem hilfreich, da auf Grund der Seltenheit von primären Knochentumoren und der damit einhergehenden schlechten Expertise vieler Radiolog\*innen eine eindeutige Befundung der primären Knochentumore eine Schwierigkeit darstellt.



## 7 Literaturverzeichnis

- Adams M, Chen W, Holcdorf D, McCusker MW, Howe PD, Gaillard F. 2019.** Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. 02 2019, Bd. 63, 1, S. 27-31.
- Balach T, Stacy GS, Peabody TD. 2011.** The clinical evaluation of bone tumors. 11 2011, Bd. 49, 6, S. 1079-1093.
- Bandyopadhyay O, Biswas A, Bhattacharya BB. 2019.** Bone-Cancer Assessment and Destruction Pattern Analysis in Long-Bone X-ray Image. 04 2019, Bd. 32, 2, S. 300-313.
- . 2019. Bone-Cancer Assessment and Destruction Pattern Analysis in Long-Bone X-ray Image. 04 2019, Bd. 32, 2, S. 300-313.
- Bayramoglu N, Tiulpin A, Hirvasniemi J, Nieminen MT, Saarakkala S. 2020.** Adaptive segmentation of knee radiographs for selecting the optimal ROI in texture analysis. 07 2020, Bd. 28, 7, S. 941-952.
- Benndorf M, Bamberg F, Jungmann PM. 2021.** The Lodwick classification for grading growth rate of lytic bone tumors: a decision tree approach. . 07 2021.
- Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, Halpern EF, Hess CP, Schiebler ML, Weiss CR. 2020.** Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers-From the Radiology Editorial Board. 03 2020, Bd. 294, 3, S. 487-489.
- Caracciolo JT, Temple HT, Letson GD, Kransdorf MJ. 2016.** A modified Lodwick-Madewell grading system for the evaluation of lytic bone lesions. 07 2016, Bd. 207, 1, S. 150-156.
- Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, Kadoury S, Tang A. 2012.** Deep Learning: A Primer for Radiologists. 12 2012, Bd. 37, 7, S. 2113-2131.
- Chauhan NK, Singh K. 2018.** A Review on Conventional Machine Learning vs Deep Learning. 2018, S. 347-352.
- Choi JH, Ro JY. 2021.** The 2020 WHO Classification of Tumors of Bone: An Updated Review. 05 2021, Bd. 28, 3, S. 119-138.
- Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, Kim JY, Moon SH, Kwon J, Lee HJ, Noh YM, Kim Y. 2018.** Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. 08 2018, Bd. 89, 4, S. 468-473.
- Costelloe CM, Madewell JE. 2013.** Radiography in the initial diagnosis of primary bone tumors. *AJR*. 01 2013, Bd. 200, 1, S. 3-7.
- Davies AM, Sundaram M, James SL. 2009.** *Imaging of Bone Tumors and Tumor-Like Lesions*. s.l. : Springer, 2009.

**Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, Behr SC, Flavell RR, Huang SY, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Hernandez Pampaloni M, Hadley D, Franc BL. 2019.** A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. 02 2019, Bd. 290, 2, S. 456-464.

**Do BH, Langlotz C, Beaulieu CF. 2017.** Bone Tumor Diagnosis Using a Naïve Bayesian Model of Demographic and Radiographic Features. 10 2017, Bd. 30, 5, S. 640-647.

**Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. 2020.** Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. 11 2020, Bd. 39, 11, S. 3679-3690.

**Erickson BJ, Korfiatis P, Akkus Z, Kline TL. 2017.** Machine Learning for Medical Imaging. 03 2017, Bd. 37, 2, S. 505-515.

**Erlemann, R. 2009.** Basisdiagnostik von Knochentumoren. *Radiologe*. 04 2009, Bd. 49, 4, S. 355-370.

**Farkaš I, Masulli P, Otte S, Wermter S. 2021.** *Artificial Neural Networks and Machine Learning – ICANN 2021*. s.l. : Springer International Publishing, 2021.

**Fletcher, CDM. 2013.** *WHO classification of tumours of soft tissue and bone*. Geneva, Switzerland : WHO, 2013. ISBN: 978-92-832-2434-1.

**Gemescu IN, Thierfelder KM, Rehnitz C, Weber MA. 2019.** Imaging Features of Bone Tumors: Conventional Radiographs and MR Imaging Correlation. 11 2019, Bd. 27, 4, S. 753-767.

**Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. 2019.** Artificial Intelligence in Musculoskeletal Imaging: Current Status and Future Directions. 09 2019, Bd. 213, 3, S. 506-513.

**Hamid, R. Nathan, T. JunYoung, G. Amir, S. Ian, R. Silvio, S. 2019.** Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, C. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 16–20 . 06 2019, S. 658–666.

**Han XG, Tian W. 2019.** Artificial intelligence in orthopedic surgery: current state and future perspective. 11 2019, Bd. 132, 21, S. 2521-2523.

**He Y, Pan I, Bao B, Halsey K, Chang M, Liu H, Peng S, Sebro RA, Guan J, Yi T, Delworth AT, Eweje F, States LJ, Zhang PJ, Zhang Z, Wu J, Peng X, Bai HX. 2020.** Deep learning-based classification of primary bone tumors on radiographs: A preliminary study. 12 2020, Bd. 62.

**Jundt, G. 2018.** Neues zur WHO-Klassifikation der Knochentumoren. *Pathologe*. 03 2018, Bd. 39, S. 107-116.

**King, BF Jr. 2017.** Guest Editorial: Discovery and Artificial Intelligence. 12 2017, Bd. 209, 6, S. 1189-1190.

**Kroque JD, Cheng KV, Hwang KM, Toogood P, Meinberg EG, Geiger EJ, Zaid M, McGill KC, Patel R, Sohn JH, Wright A, Darger BF, Padrez KA, Ozhinsky E, Majumdar S, Pedoia V. 2020.** Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. 03 2020, Bd. 2, 2.

—. **2020.** Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. 03 2020, Bd. 2, 2.

**Lakhani P, Prater AB, Hutson RK, Andriole KP, Dreyer KJ, Morey J, Prevedello LM, Clark TJ, Geis JR, Itri JN, Hawkins CM. 2018.** Machine Learning in Radiology: Applications Beyond Image Interpretation. 02 2018, Bd. 15, 2, S. 350-359.

**Lakhani P, Sundaram B. 2017.** Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. 08 2017, Bd. 284, 2, S. 574-582.

**Lalam R, Bloem JL, Noebauer-Huhmann IM, Wörtler K, Tagliafico A, Vanhoenacker F, Nikodinovska VV, Sanal HT, Woude HV, Papakonstantinou O, Åström G, Davies M, Isaac A, Weber MA. 2017.** ESSR Consensus Document for Detection, Characterization, and Referral Pathway for Tumors and Tumorlike Lesions of Bone. *Semin Musculoskelet Radiol.* 11 2017, Bd. 21, 5, S. 630-647.

**LeCun Y, Bengio Y, Hinton G. 2015.** Deep learning. 05 2015, Bd. 521, 7553, S. 436-444.

**Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, Cho K, Chang G, Deniz CM. 2020.** Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. 09 2020, Bd. 296, 3, S. 584-593.

**Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P. 2014.** Microsoft COCO: Common Objects in Context. 2014.

**Lodwick GS, Wilson AJ, Farrell C, Virtama P, Dittrich F. 1980.** Determining growth rates of focal lesions of bone from radiographs. 03 1980, Bd. 134, 3, S. 577–583.

**Lodwick, GS. 1964.** Radiographic diagnosis and grading of bone tumors, with comments on computer evaluation. *Proc Natl Cancer Conf.* 1964, Bd. 5, S. 369–380.

**Macedo F, Ladeira K, Pinho F, Saraiva N, Bonito N, Pinto L, Goncalves F. 2017.** Bone Metastases: An Overview. 05 2017, Bd. 11, 1.

**Madewell JE, Ragsdale BD, Sweet DE. 1981.** Radiologic and pathologic analysis of solitary bone lesions. Part I: internal margins. *Radiol Clin North Am.* 12 1981, Bd. 19, 4, S. 715-748.

**McHugh ML, . 2012.** Interrater reliability: the kappa statistic. 2012, Bd. 22, 3, S. 276-282.

- Miller, TT. 2008.** Bone tumors and tumorlike conditions: analysis with conventional radiography. *Radiology*. 03 2008, Bd. 246, 3, S. 662-74.
- Mintz DN, Hwang S. 2014.** Bone tumor imaging, then and now: review article. *HSS J*. 10 2014, Bd. 10, 3, S. 230-239.
- Motamedi K, Seeger LL. 2011.** Benign bone tumors. 11 2011, Bd. 49, 6, S. 1115-1134.
- Mundy, GR. 2002.** Metastasis to bone: causes, consequences and therapeutic opportunities. 08 2002, Bd. 2, 8, S. 584-593.
- Nascimento D, Suchard G, Hatem M, de Abreu A. 2014.** The role of magnetic resonance imaging in the evaluation of bone tumours and tumour-like lesions. 08 2014, Bd. 5, 4, S. 419-440.
- Pan SJ, Yang Q. 2010.** A Survey on Transfer Learning. 10 2010, Bd. 22, 10, S. 1345-1359.
- Park SH, Han K. 2018.** Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. 03 2018, Bd. 286, 3, S. 800-809.
- Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. 2017.** Automatic differentiation in PyTorch. [Online] 28. 10 2017. [https:// openreview.net/forum?id=BJJsrnCZ..](https://openreview.net/forum?id=BJJsrnCZ..)
- Pedregosa, Fabian, et al. 2011.** Scikit-learn: Machine Learning in Python. 10 2011, Bd. 12, 85, S. 2825–2830.
- Pesapane F, Codari M, Sardanelli F. 2018.** Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. 10 2018, Bd. 2, 1.
- Pfeiffer D, Pfeiffer F, Rummeny E. 2020.** Advanced X-ray Imaging Technology. 2020, Bd. 216, S. 3-30.
- Poole DI, Goebel RG, Mackworth AK. 1998.** *Computational Intelligence: A Logical Approach*. s.l. : Oxford University Press, 1998. 978-0195102703.
- Powers, DMW. 2011.** Evaluation: from precision, recall and F-measure to ROC, informed- ness, markedness & correlation. 2011, Bd. 2, 1, S. 37-63.
- Ragsdale BD, Madewell JE, Sweet DE. 1981.** Radiologic and pathologic analysis of solitary bone lesions. Part II: periosteal reactions. 12 1981, Bd. 19, 4, S. 749-783.
- Redondo A, Bagué S, Bernabeu D, Ortiz-Cruz E, Valverde C, Alvarez R, Martinez-Trufero J, Lopez-Martin JA, Correa R, Cruz J, Lopez-Pousa A, Santos A, García Del Muro X, Martin-Broto J. 2017.** Malignant bone tumors (other than Ewing's): clinical practice guidelines for diagnosis, treatment and follow-up by Spanish Group for Research on Sarcomas (GEIS). 12 2017, Bd. 80, 6, S. 1113-1131.
- Reinus WR, Wilson AJ, Kalman B, Kwasny S. 1994.** Diagnosis of focal bone lesions using neural networks. 06 1994, 6, S. 606-611.

- Robert Koch Institute and the Association of Population-based, Cancer Registries in Germany. 2018.** Cancer in Germany in 2013/2014. 11th edition. Berlin : s.n., 2018. ISBN: 978-3-89606-289-5.
- Siegel RL, Miller KD, Jemal A. 2019.** Cancer statistics, 2019. 01 2019, Bd. 69, 1, S. 7-34.
- Stiller CA, Trama A, Serraino D, Rossi S, Navarro C, Chirilaque MD, Casali PG and Group, RARECARE Working. 2013.** Descriptive epidemiology of sarcomas in Europe: report from the RARECARE project. *Eur J Cancer*. 02 2013, Vol. 49, 03, pp. 684-695.
- Twisk, JWR. 2003.** *Applied Longitudinal Data Analysis for Epidemiology A Practical Guide*. s.l. : Cambridge University Press, 2003. 978-0521525800.
- Uhl M, Hergert GW. 2015.** *Radiologische Diagnostik von Knochentumoren*. s.l. : Thieme, 2015. doi: 10.1055/b-003-127162.
- Uhl M, Herget G, Baumhoer D, Ruf J. 2022.** *Bildgebende Diagnostik und ihre Aufgaben*. s.l. : Thieme, 2022. doi:10.1055/b000000484.
- Unni, K. Krishnan Inwards, Carrie Y. 2010.** *Dahlin's Bone Tumors*. [Hrsg.] Lippincott Williams & Wilkins. 2010. Bd. 6. Edition.
- Vogrin M, Trojner T, Kelc R. 2020.** Artificial intelligence in musculoskeletal oncological radiology. 11 2020, Bd. 55, 1, S. 1-6.
- von Schacky CE, Sohn JH, Liu F, Ozhinsky E, Jungmann PM, Nardo L, Posadzy M, Foreman SC, Nevitt MC, Link TM, Padoia V. 2020.** Development and Validation of a Multitask Deep Learning Model for Severity Grading of Hip Osteoarthritis Features on Radiographs. 04 2020, Bd. 295, 1, S. 136-145.
- von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Gassert FG, Foreman SC, Gassert FT, Jung M, Jungmann PM, Russe MF, Mogler C, Knebel C, von Eisenhart-Rothe R, Makowski MR, Woertler K, Burgkart R, Gersing AS. 2021.** Multitask Deep Learning for Segmentation and Classification of Primary Bone Tumors on Radiographs. *Radiology*. 11 2021, 301(2), S. 398-406.
- Wu Y, Kirillov A, Massa F, Lo WY, Girshick R. 2019.** Detectron2. *GitHub Web site*. [Online] 2019. <https://github.com/facebookresearch/detectron2>.

## 8 Abbildungsverzeichnis

Abbildung 1: Typische topografische Lokalisation von Knochentumoren (Uhl M, 2015). .....	14
Abbildung 2: Schematische Darstellung der Knochenosteolysen, in Anlehnung an (Lalam R, 2017).....	16
Abbildung 3: Übersicht der Periostreaktionen, in Anlehnung an (Lalam R, 2017) ....	17
Abbildung 4: Entscheidungsbaum zur Klassifizierung nach Lodwick Grad I-III (Lodwick GS, 1980).....	19
Abbildung 5: Künstliche Intelligenz und die Unterkategorien Machine learning und Deep learning .....	20
Abbildung 6: Visualisierung der drei Schichten der neuronalen Netzwerke .....	21
Abbildung 7: Flussdiagramm der Patientenauswahl .....	25
Abbildung 8: Darstellung der Segmentierung.....	27
Abbildung 9: Vereinfachte Darstellung der Modell-Architektur (von Schacky CE, 2021) .....	29
Abbildung 10: Schematische Visualisierung der Intersection over Union Kennzahl .	30
Abbildung 11: Konfusionsmatrix der internen Testreihe. Die Zahlen entsprechen den verschiedenen Vorhersagen (von Schacky CE, 2021).....	34
Abbildung 12: Konfusionsmatrix der externen Testreihe. Die Zahlen entsprechen den verschiedenen Vorhersagen (von Schacky CE, 2021).....	35
Abbildung 13: Beispiel für die korrekte Bewertung als Chondrosarkom (von Schacky CE, 2021) .....	39
Abbildung 14: Beispiel für die korrekte Bewertung als NOF (von Schacky CE, 2021) .....	40
Abbildung 15: Beispiel für die korrekte Klassifikation als benigner Tumor (von Schacky CE, 2021).....	41
Abbildung 16: Beispiel für die korrekte Klassifikation als maligner Tumor (von Schacky CE, 2021).....	42
Abbildung 17: Beispiel für eine fehlerhafte Bewertung (orange Segmentierungs-Maske). Die eigentliche Läsion befindet sich im Trochanter major links/proximalen Femur links (lila Segmentierungs-Maske) (von Schacky CE, 2021) .....	43

## 9 Tabellenverzeichnis

Tabelle 1: Einteilung der Knochentumore nach WHO 2020 in Anlehnung an (Choi JH, 2021). .....	10
Tabelle 2: Lokalisation der Knochentumore im Skelett, in Anlehnung an (Unni, 2010). .....	11
Tabelle 3: Typisches Manifestationsalter von Knochentumoren, in Anlehnung an (Lalam R, 2017). .....	12
Tabelle 4: Übersicht der Einteilung nach Lodwick Grad I-III (Lodwick GS, 1980) .....	18
Tabelle 5: Verteilung der Patient*innen .....	32
Tabelle 6: Fortsetzung der Tabelle 5: Verteilung der Patient*innen .....	33
Tabelle 7: Leistung des DL-Modells bei der Identifizierung maligner Tumore in den internen und externen Testgruppen sowie bei der Klassifizierung von Tumorentitäten, der Platzierung von Bounding Boxen und der Segmentierung (von Schacky CE, 2021). .....	36
Tabelle 8: Leistung des DL-Modells, von Assistenzärzt*innen für Radiologie und von Fachärzt*innen für Radiologie mit muskuloskelettalem Schwerpunkt bei der Identifizierung maligner Tumoren und der Klassifizierung von Tumorentitäten in der externen Testgruppe (von Schacky CE, 2021).....	38

## 10 Danksagung

Die vorliegende Dissertation schließt meine Arbeit als Doktorandin am Institut für diagnostische und interventionelle Radiologie der Technischen Universität München am Graduate Center of Medicine and Health der Technischen Universität München von Mai 2020 bis Dezember 2023 ab.

An erster Stelle möchte ich mich bei Frau Prof. Dr. med. Alexandra Gersing für ihre kontinuierliche Unterstützung, ihr umfassendes Interesse, ihre wissenschaftliche Begleitung und ihr Vertrauen in meine Arbeit herzlich bedanken.

Mein aufrichtiger Dank gilt ebenfalls Herrn Dr. med. Claudio von Schacky für seine großartige Unterstützung, sein Coaching, sein Interesse und sein Vertrauen in meine Arbeit.

Zudem möchte ich mich bei Frau PD Dr. Susanne Bensler und Frau Prof. Dr. Pia Baumann-Jungmann für ihre Ratschläge und Diskussionen bei der wissenschaftlichen Arbeit bedanken.

Zuletzt möchte ich mich bei meiner ganzen Familie für ihre grenzenlose und ausdauernde Unterstützung und bedingungslose Liebe bedanken. Ein besonderer Dank geht dabei an meinen Mann Benjamin für seine stetige, selbstlose und liebevolle Unterstützung und Geduld.

Valerie Sophie Schäfer

Kirchheim unter Teck, Dezember 2023



## 11 Veröffentlichungen

**von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Jung M, Jungmann PM, Russe MF, Foreman SC, Gassert FG, Gassert FT, Schwaiger BJ, Mogler C, Knebel C, von Eisenhart-Rothe R, Makowski MR, Woertler K, Burgkart R, Gersing AS. 2022.** Development and evaluation of machine learning models based on X-ray radiomics for the classification and differentiation of malignant and benign bone tumors. *Eur Radiol.* 2022 Sep, 32(9), S. 6247-6257

**von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Gassert FG, Foreman SC, Gassert FT, Jung M, Jungmann PM, Russe MF, Mogler C, Knebel C, von Eisenhart-Rothe R, Makowski MR, Woertler K, Burgkart R, Gersing AS. 2021.** Multitask Deep Learning for Segmentation and Classification of Primary Bone Tumors on Radiographs. *Radiology.* 11 2021, 301(2), S. 398-406.