Thesis

Differentiating childhood benign and malignant bone lesions in radiological examination using artificial intelligence

submitted by

Evelyn Kupsch

in partial fulfillment of the requirements for the degree of

Doktorin der gesamten Heilkunde

(Dr^{in}. med. univ.)

at the

Medical University of Graz

executed at the

Department of Radiology

Division of Pediatric Radiology

under the supervision of

Univ.-Prof. Priv.-Doz. Dr.med.univ. Dr.scient.med. Sebastian Tschauner

Dr. Anja Dutschke

Graz, date 09.09.2025

Declaration of Academic Integrity

I hereby confirm that the present diploma thesis is the result of my own independent scholarly work. I also confirm that in all cases, where material from the work of others (in books, articles, essays, dissertations, and on the internet) is acknowledged, quotations and paraphrases are clearly indicated. No material other than that cited in the reference list has been used. I have read and understood the Medical University's regulations and procedures concerning plagiarism.

Furthermore, I hereby declare that if artificial intelligence (AI) tools were used for the generation and/or correction of certain text passages in the creation of this work, such employment was conducted in compliance with ethical principles, academic integrity, and the regulations of my university. Additionally, it was ensured that this usage was transparently disclosed and appropriately attributed.

Graz, 09.09.2025 *Evelyn Kupsch m.p.*

I. Acknowledgements

First and foremost, I would like to sincerely thank my supervisor Univ.-Prof. Priv.-Doz. Dr.med.univ. Dr.scient.med. Sebastian Tschauner, for his support, patience, and constant availability for questions of all kinds. His support and expert guidance were essential in shaping and completing this work. I am grateful for the opportunity to learn from him and for his encouragement during challenging moments.

I am also forever grateful to my parents, Anna and Walter, for making my medical education possible. Without their support, love and belief in me, none of this would have been achievable.

Last but not least, I want to thank my fiancé Stefan, whose encouragement and patience kept me going through late nights and stressful days. His unwavering support has been a constant comfort, and I couldn’t have done this without him.

II. Zusammenfassung in Deutsch

Einleitung: Primäre Knochentumore im Kindes- und Jugendalter sind selten und stellen daher eine diagnostische Herausforderung dar. Maligne Läsionen sind schwerwiegende Diagnosen, die eine Reihe von Therapien erfordern und einen letalen Verlauf haben können. Eine frühzeitige Erkennung und zuverlässige Einschätzung der Dignität solcher Läsionen ist von hoher Relevanz. Ziel dieser Diplomarbeit war es, einen Datensatz mit Röntgenbilder pädiatrischer maligner und benigner Knochenläsionen zu erstellen und diesen zur Evaluierung von EfficientNet-Modellen B0 – B7 hinsichtlich ihrer Fähigkeit zur Dignitätsklassifikation zu verwenden.

Patient*innen und Methoden: In die Studie eingeschlossen wurden Patient*innen, die zum Untersuchungszeitpunkt zwischen 0 – 19 Jahre alt waren, mit einem Untersuchungszeitraum zwischen dem 01.01.2004 bis zum 30.05.2024. Die Diagnosen wurden entweder histopathologisch oder, sofern keine Biopsie notwendig war oder vorlag, durch ihr typisches radiologisches Erscheinungsbild bestätigt. Um Störfaktoren zu minimieren, wurden nur präinterventionelle Bilder eingeschlossen und solche nach Biopsien, Operationen, Chemotherapie und Radiotherapie ausgeschlossen. Der finale Datensatz umfasste 800 Röntgenbilder von 228 individuellen Patient*innen. Dieser wurde genutzt, um EfficientNet Varianten B0 – B7 zu trainieren und anschließend zu testen. Als Testparameter wurden Genauigkeit (Accuracy), Sensitivität (Recall), positive Vorhersagewert (Precision) und F1-Score verwendet. Zudem wurden Precision-Recall-Kurven (PR-Kurven) und Receiver-Operating-Characteristic-Kurven (ROC-Kurven) erstellt und deren Fläche unter der Kurve (AUC) berechnet.

Ergebnisse: Das Alter der Patient*innen lag zwischen 0.10 und 17.60 Jahren. Es handelte sich um 396 Bilder (38.25%) weiblicher Patientinnen und 464 Bilder (61.75%) männlicher Patienten. Der Datensatz umfasste 707 benigne (88.37%) und 93 maligne (11.63%) Läsionen. Alle Efficient-Net-Varianten erzielten bei benignen Tumoren eine sehr hohe Sensitivität (0.979 – 1.000) und einen guten positive Vorhersagewert (0.909 – 0.914). Der höchste F1-Score wurde mit 0.954 bei EfficientNet-B2 erreicht. Die Erkennung maligner Läsionen war deutlich schwächer. Die Sensitivität lag zwischen 0.237 und 0.301, der positive

Vorhersagewert variierte stark zwischen 0.651 und 1.000. Die F1-Scores fielen dementsprechend niedrig aus (0.379 – 0.424). Bezogen auf die Gesamtheit aller Fälle zeigte EfficientNet-B2 die beste Performance mit einer Genauigkeit von 0.915 und einem F1-Score von 0.689. Die höchste PR-AUC erzielte B3 (0.624), dicht gefolgt von B1, B2 und B5.

Schlussfolgerung: EfficientNet-Modelle sind gut geeignet zur Erkennung benigner Knochentumor im Kindsalter, gezeigt durch die hohe Sensitivität in dieser Gruppe. Die Zuverlässigkeit bei malignen Läsionen war hingegen unzureichend, mit einer durchschnittlichen Sensitivität von 27%. Die Modelle identifizierten maligne Läsionen nur dann korrekt, wenn eine hohe Klassifikationssicherheit vorlag, was in einer niedrigen Falsch-Positiv-Rate, aber einer hohen Falsch-Negativ-Rate resultierte. Für diese Ergebnisse dürfte hauptsächlich das starke Klassenungleichgewicht im Datensatz ursächlich sein (88% benigne vs. 12% maligne Läsionen). Dies erschwert die Generalisierungsfähigkeit der Modelle erheblich. Für eine zuverlässige Anwendung in der klinischen Praxis ist ein Verbesserung der Modellleistung bei malignen Läsionen notwendig. Dies kann durch eine Erweiterung des Datensatzes, beispielweise durch synthetische Datengenerierung oder Datenaugmentation, erreicht werden.

III. Abstract in Englisch

Background: Primary bone tumors in children and adolescents are rare, therefore posing a significant diagnostic challenge. Malignant lesions are critical diagnoses that require intensive treatment and can be fatal. Early detection and reliable assessment of the lesion's nature are of great clinical importance. The aim of this thesis was to create a dataset of radiographic images of pediatric benign and malignant bone lesions and to use it to evaluate the ability of EfficientNet models (B0 – B7) in classifying lesion dignity.

Patients and Methods: The study included patients aged 0 to 19 years who underwent radiographical examination between 01.01.2004 and 30.05.2024. Diagnoses were either confirmed histopathologically or, if biopsy was not necessary on unavailable, made based on characteristic radiological appearance. To minimize confounding factors, only pre-interventional images were included, and those taken after biopsy, surgery, chemotherapy, or radiotherapy were excluded. The final dataset consisted of 800 X-ray images from 228 individual patients. These images were used to train and test EfficientNet variants B0 to B7. Evaluation metrics included accuracy, recall, precision, and F1-score. Additionally, Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves were created and the area under the curve (AUC) was calculated.

Results: The patients' ages ranged from 0.10 to 17.60 years. The dataset included 396 images (38.25%) from female and 464 images (61.75%) from male patients. Of the total, 707 lesions (88.37%) were benign and 93 (11.63%) were malignant. All EfficientNet models demonstrated very high recall (0.979 – 1.000) and good precision (0.909 – 0.914) in detecting benign bone tumors. The highest F1-score for benign lesions was achieved by EffiecientNet-B2 (0.954). The performance on malignant lesions was significantly lower. Recall ranged between 0.237 to 0.301, with precision values varying between 0.651 and 1.000. The corresponding F1-scores were low (0.379 – 0.424). Overall, EfficientNet-B2 showed the best performance across all cases, with an accuracy of 0.915 and an F1-score of 0.689. The highest PR-AUC was achieved by B3 (0.624), closely followed by B1, B2, and B5.

Discussion: EfficientNet models are well suited for detecting benign pediatric bone tumors, as demonstrated by their high recall in this particular group. However, reliability for malignant lesions was inadequate, with an average recall of only 27%. The models correctly identified malignant lesions only when classification certainty was high, resulting in low false-positive rate but a high false-negative rate. The main reason for these results is likely due to the significant class imbalance in the dataset (88% benign vs. 12% malignant), which greatly limits model generalizability. For reliable application in clinical practice, the model performance for malignant cases must be improved. This can be achieved by expanding the dataset, for instance through synthetic data generation or data augmentation.

Table of contents

IV. List of abbreviations

5-YSR	5-Year Survival Rate
18F-FDG PET/CT	18F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography
ABC	Aneurysmal Bone Cyst
AI	Artificial Intelligence
AJCC	American Joint Committee on Cancer
ALARA	As Low As Reasonably Achievable
AUC	Area Under the Curve
CAM	Class Activation Mapping
CE	Conformité Européenne
CBL	Chondroblastoma
CMF	Chondromyxoid Fibroma
CNN	Convolutional Neural Network
CNS	Central Nervous System
CT	Computed Tomography
CV	Computer Vision
DICOM	Digital Imaging and Communications in Medicine Format
DL	Deep Learning
ENC	Enchondroma
ES	Ewing Sarcoma

ESFT	Ewing Sarcoma Family of Tumors
FCD	Fibrous Cortical Defect
FD	Fibrous Dysplasia
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
GCTB	Giant Cell Tumor of Bone
HME	Hereditary Multiple Exostoses
ID	Identification
LCH	Langerhans Cell Histiocytosis
La	Locally Aggressive
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NOF	Non-Ossifying Fibroma
NOS	Not Otherwise Specified
NPV	Negative Predictive Value
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OBL	Osteoblastoma
OCH	Osteochondroma
OO	Osteoid Osteoma
OS	Osteosarcoma

PACS	Picture Archiving and Communication System
PBT	Primary Bone Tumors
PPV	Positive Predictive Value
PR	Precision-Recall
Rm	Rarely Metastasizing
ROC	Receiver Operating Characteristic
SBC	Simple Bone Cyst
SHA-3	Secure Hash Algorithm 3
Tc-99m	Technetium-99m
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
UTC	Coordinated Universal Time
WHO	World Health Organization
XAI	Explainable Artificial Intelligence

V. List of figures

VI. List of tables

1. Introduction

The diagnostic imaging of benign and malignant bone tumors in children has always been particularly challenging. The differential diagnosis is based on various parameters, including the patient's age, tumor location, and image morphology using different radiological modalities, such as ultrasound, plain radiography, computed tomography (CT), and magnetic resonance imaging (MRI). In some cases, nuclear medicine procedures are also employed. Often, a definitive diagnosis requires a tissue sample, for which imaging is essential to perform a targeted biopsy. One of the key diagnostic issues is distinguishing between benign and malignant bone lesions. (1) Malignant tumors require immediate and often complex treatment, making early detection all the more important. More precise imaging could potentially help in avoiding unnecessary invasive procedures in benign lesions.

Malignant primary bone tumors (PBTs), namely osteosarcoma and Ewing sarcoma, are rare in children. Neoplasm of the bone takes sixth place in the incidence ranking of pediatric cancers but takes third place in cause of cancer deaths in children and adolescents. (2) Due to their rarity, it can be difficult to classify a lesion quickly and correctly outside of specialized centers. This can have serious consequences for necessary early treatment. Aside from benign lesions, other differential diagnoses include acute or chronic infections. Early symptoms, such as pain or swelling, are often nonspecific, and imaging does not always show clear results. X-ray imaging usually serves as a starting point and can be followed up by further examinations.

In recent years, the use of artificial intelligence (AI) in medicine has become increasingly important. Deep learning (DL) in particular shows great potential in radiological image analysis. In the future, radiologists might be supported by AI-based tools, especially when interpreting complex image patterns, such as those found in bone tumors. A high-quality data set on which a neural network can be trained to distinguish between benign and malignant entities is a prerequisite for developing such a system. These automated systems could provide valuable assistance in everyday clinical practice and contribute to decision support. This

could streamline assessment, particularly outside specialized centers, and promote faster referrals.

This thesis aims to compile a usable dataset of pediatric bone tumors consisting of benign and malignant lesions. Conventional radiography was selected as a modality, because it is widely available and often the first line of diagnostics. The database forms the basis for developing and evaluating an AI-supported classification system for pediatric PBTs. Due to the limited data available in the literature and the lack of established decision support systems in this field, this study aims to contribute to the development of a practical tool.

Following this introduction, the medical background of childhood bone tumors will be outlined in *Chapter 2. Medical background*, including epidemiology, classification, and short descriptions of the different entities. In the subsequent *Chapter 3. Diagnostic imaging of bone lesions*, the current diagnostic pathway and imaging characteristic will be explained. *Chapter 4. Artificial intelligence in medical imaging* contains an introduction to AI in medicine, as well as a literature review of current applications in musculoskeletal oncological radiology. After establishing this foundation, the materials and methodology of this thesis will be described in *Chapter 5. Materials and methods*. Lastly the results will be presented in *Chapter 6. Results* and discussed in *Chapter 7. Discussion*.

2. Medical background

2.1 Primary bone tumors (PBTs) in children

2.1.1 Epidemiology

Malignant diseases in children, adolescents and young adults are rare overall, with an incidence of 155.8 per million children. Different types of cancer present at different frequencies in different age groups, with leukemias, lymphomas and central nervous system (CNS) tumors being the most common across all age groups. Malignant bone tumors account for 4.7% of all tumors in 0 – 14-year-olds and 7.8% in 15 – 19-year-olds, placing them in sixth position of cancer ranking in both age groups. (3)

Current American epidemiological data summarized by *Siegel et al.* (2) estimates the annual incidence of all bone tumors in 0 – 14-year-olds to be around 7.3 per million, while in the 15 – 19-year-old group it is 14.3 per million.

Osteosarcoma (OS) is the most common malignant tumor in minors. In children under 14 years of age the incidence is 4.1 per million, with a 5-year survival rate (5-YSR) of 65%. In the age group from 15 – 19 the incidence increases to 7.9 million, with the survival rate remaining at around 64%. (2)

Ewing sarcoma (ES) is the second most common malignant bone tumor in minors. In children under the age of 14 the incidence is 2.4 per million with a 5-YSR of 81%. Among adolescents between 15 – 19 years of age, the incidence rises to 4.3 per million while the survival rate is lower at 68%. (2)

This shows that the mortality rate for malignant bone tumors is comparatively high in relation to the low frequency of the disease, ranking third among death caused of cancer in children and adolescents aged 0 – 19 years, following CNS tumors and leukemia. (2)

Data from the German Childhood Cancer Registry paint a similar picture, although it only includes children up to the age of 15. Malignant bone tumors account for 3 – 5% of pediatric oncological cases. The incidence rate is 5.6 per million male children and 5.4 per million female children. (4) The male-to-female ratio is 1.3 : 1 in OS (5), and 1.5 : 1 in ES (6).

Compared to malignant bone tumors, benign bone tumors occur much more frequently, while still being very rare in the overall population. A recent German study by *Breden et al.* (7) examined a collective of 420 patients aged 0 – 18 years who were diagnosed with a bone lesion between the years 2003 to 2023 and found that 80% (335 patients) of the tumors were benign and 20% (85 patients) were malignant. When it came to gender, benign tumors occurred much more often in male patients (65% in male vs. 35% in female patients), while malignant tumors showed a lower male predominance (53% in male vs. 47% in female patients). It is important to note that this study was conducted at a tertiary center and only included patients who underwent surgery, radiotherapy or histological confirmation of the diagnosis. As many benign tumors do not require treatment or

are asymptomatic, thus remaining clinically undetected, it can be assumed that benign lesion occur even more frequently than shown here.

Benign tumors also demonstrate a typical distribution pattern across different age groups. A detailed overview of this can be found in chapter *2.3.1 Epidemiology of benign tumors* in *Table 2*.

2.1.2 Etiology and pathogenesis

In most cases the exact cause of benign and malignant PBTs is unclear. It has been suggested that there is a possible connection to increased bone growth rates and remodeling processes in children, as well as bone marrow distribution and maturation from yellow to red bone marrow. (7–10)

Some risk factors associated with the development of malignant bone tumors have been identified. These include age, gender, socioeconomic status, height, genetic predisposition, previous oncological therapies such as radiotherapy or pre-existing bone diseases such as Paget's disease. (11,12)

As described, OS shows a peak incidence during the second decade of life and in late adulthood. Male patients are more likely to develop OS, as well as patients who are 'taller-than-average'. A lower socio-economic status also seems to increase the risk of cancer development. (11,13) Genetic predispositions are mutations in the TP53 gene, as found in Li-Fraumeni syndrome or mutations in the RB1 gene, as well as Rothmund-Thomson syndrome. (11)

In ES, 85% of all cases show a chromosomal translocation between chromosomes 11 and 22, namely t(11;22)(q24;q12), resulting in an EWS-FLI1 gene fusion. In the remaining cases, there is fusion with another ETS transcriptions factor such as ERG, ETV1 or E1AF. (14)

The cause for the development of benign bone tumors is unclear in many cases. Some benign lesions show fundamental genetic alterations, confirming them as true neoplasms. For instance, chondroblastoma (CBL) frequently exhibits a mutation in the H3F3B gene, which codes for a variant of histone H3. (15)

Enchondroma (ECH), especially in cases of enchondromatosis like Ollier disease and Maffucci syndrome, frequently exhibit mutations in IDH1 and IDH2 genes. (16) Other tumors, such as non-ossifying fibroma (NOF) and Langerhans cell histiocytosis (LCH) show alterations in the MAPK-signaling pathway. (15,17) Giant cell tumors of bone (GCTB) are one of the few entities found more frequently in female patients, and they may be connected to hormones and pregnancy. (18) An activating GNAS mutation can often be detected in fibrous dysplasia (FD), which can be helpful in differentiating it from low-grade osteosarcoma. (15)

2.1.3 Clinical symptoms and locations

The clinical presentation of PBTs in children and adolescents is often non-specific and can initially resemble a variety of other musculoskeletal issues.

In most cases of malignant tumors, the main symptom is persistent, progressively developing pain. This pain is typically described as dull and does not respond to conventional analgesics like non-steroidal anti-inflammatory drugs (NSAIDs). A notable diagnostic warning sign is pain which worsens at night, which should be evaluated as a potential indication of malignant disease (‘red flag’). (19) However, benign tumors such as osteoid osteoma (OO) can also present as dull, nocturnal pain, while being responsive to NSAIDs. Osteoblastoma (OBL) also present with dull pain and are unresponsive to NSAIDs, but do not tend to get worse during nighttime. (8,18)

Furthermore, the presence of a palpable mass, possibly accompanied by local tenderness, may indicate a malignant process. Systemic symptoms such as lethargy, malaise and fever may also occur, although they are less specific and are particularly common in advanced disease. (19)

Another possible clinical presentation is the pathological fracture. They are the result from the structural weakening of the bone by the tumor and can be accompanied by acute pain and sudden loss of function. The initial trauma is often times inadequate. (19) Pathological fractures can occur in both malignant and large benign tumors. Classic examples are simple bone cysts (SBC), aneurysmal bone cysts (ABC), non-ossifying fibroma (NOF) and fibrous dysplasia (FD). (20)

Most benign tumors are completely asymptomatic and are often discovered incidentally as part of another investigation, for instance following trauma. (8) Additional clinical features are mentioned in the subchapters of the respective entities.

The majority of bone tumors manifest in the long tubular bones. The bones of the lower extremities are most frequently affected, especially around the knee. Short or flat bones are affected less often. The most important differences in localization are shown in *Table 3* in chapter *3.2.1 Location.*

2.1.4 Classification of bone lesions

Bone tumors arise from mesenchymal cells and can be categorized as benign, malignant or intermediate according to their biological behavior. Intermediate tumors can further be subdivided into two categories: rarely metastasizing and locally aggressive. (21–23)

Benign tumors show low recurrence tendencies, are locally confined and are not metastatic. Malignant tumors have a high potential for recurrence, are locally destructive and have a metastatic risk between 20 – 100 %. Intermediate tumors (locally aggressive) demonstrate a high potential for recurrence without metastatic tendencies, whereas intermediate tumors (rarely metastasizing) show a risk for metastasis of above 2%. (23)

Additionally, ‘tumor-like lesions’ have been described as appearing like tumors on imaging but do not show neoplastic cell growth. This subgroup has been mostly omitted from the latest World Health Organization (WHO) classification. This is due to the discovery of genetic alterations in tumors previously thought to be developmental disorders in recent years, making the distinction between benign and tumor-like lesions less clear cut. (24)

According to the latest WHO classification, namely the *5th edition of the WHO Classification of Tumors of Soft Tissue and Bone 2020* (21), eight different subgroups are defined based on their histological growth pattern. These include (1) chondrogenic, (2) osteogenic, (3) fibrogenic, (4) vascular, (5) osteoclastic giant

cell-rich, (6) notochordal, (7) other mesenchymal and (8) hematopoietic tumors of bone.

Ewing sarcoma is newly listed in a separate chapter called *Undifferentiated Small Round Cell Sarcomas of Bone and Soft Tissue* together with other tumors of the Ewing Sarcoma Family of Tumors (ESFT) to show that these tumors are of similar histology but can develop in vastly different locations in both bone and soft tissue. (21)

Table 1 shows a selection of tumors relevant to this thesis grouped by their histology and biological behavior.

	(1) Osteogenic	(2) Chondrogenic	(5) Osteoclastic giant cell-rich	(7) Other mesenchymal tumors	(8) Hematopoietic	Undifferentiated small round cell sarcomas of soft tissue and bone:
Benign	OO	OCH CBL CMF ENC	ABC	NOF SBC FD	Benign form of LCH	
Intermediate	OBL (*la*)		GCTB (*la* + *rm*)			
Malignant	OS					ES

Table 1 – Classification of bone tumors by their histology and biological behavior modified after (21)
OO – Osteoid osteoma, OBL – Osteoblastoma, OS – Osteosarcoma, OCH – Osteochondroma, CBL – Chondroblastoma, CMF – Chondromyxoid fibroma, ENC – Enchondroma, ABC – Aneurysmal bone cyst, GCTB – Giant cell tumor of bone, NOF – non-ossifying fibroma, SBC – Simple bone cyst, FD – fibrous dysplasia, LCH – Langerhans cell histiocytosis, ES – Ewing sarcoma; la – locally aggressive, rm – rarely metastasizing

In addition to the WHO 2020 classification, there is also the *WHO 2022 Classification of Pediatric Tumors* (25). The latter is largely based on the former but shows some differences. In the WHO 2022 classification, bone tumors are

divided into three subgroups. Group (1) osteogenic and (2) chondrogenic remain mostly unchanged, group (5) to (7) are combined under the title *Others*. Ewing sarcoma once again falls into its own subgroup, and Langerhans cell histiocytosis, shown in the WHO 2020 as group (8) hematopoietic bone tumors, is found in its own subchapter of hematolymphoid disorders rather than directly in the bone tumor subchapter. (25,26)

All relevant entities are described in chapters *2.2 Malignant bone tumors* and *2.3 Benign bone tumors.*

2.1.5 Staging of bone tumors

The Enneking and the AJCC (American Joint Committee on Cancer) staging systems are two establishes approaches for classifying and staging bone tumors

Described by *Enneking et al*. (27) in 1980, the Enneking classification is a widely used system in orthopedic oncology suitable for both benign and malignant bone lesions. It is based on the tumor's biological behavior, its growth pattern inside and outside the bone, and its anatomical relationship to the surrounding area.

Benign tumors are divided into three stages (28):

> Stage 1 – latent/ inactive: These are not actively growing, asymptomatic lesions, that are usually discovered incidentally. They are well defined, often with sclerotic margins.
>
> Stage 2 – active: These tumors show slow but progressive growth, are usually confined to the bone and can lead to symptoms such as pain or pathological fracture. The margins are well defined but without a sclerotic rim.
>
> Stage 3 – aggressive/ locally invasive: These lesions show destructive growth with possible cortical breakthroughs and involvement of surrounding soft tissue.

Malignant bone tumors are described on the basis of their grade (G1 – low grade, G2 – high grade), tumor spread (T1 – intracompartmental, T2 –

extracompartmental) and metastases (M0 – no metastases, M1 – distant metastases). This results in the following stages (27):

I: G1 M0 (IA: T1, IB: T2)

II: G2 M0 (IIA: T1, IIB: T2)

III: metastatic disease, regardless of grade or tumor spread

The classification has direct therapeutic relevance, as it serves as a basis for surgical planning, in particular with regard to the required resection extent. (27)

The Enneking classification also forms the basis of the current AJCC classification system, which is used for staging, overall oncological assessment and estimation of prognosis. (29)

2.2 Malignant bone tumors

2.2.1 Osteosarcoma

Osteosarcoma (OS) is the most common malignant PBT in children and adolescents. It originates from primitive mesenchymal cells that produce osteoid and/ or immature bone. Most cases develop in bone tissue, although rare instances of soft tissue osteosarcoma have been described. (30)

OS typically presents with non-specific symptoms, such as dull, load-dependent pain that may be exacerbated by physical activity or trauma. As the disease progresses, swelling, a palpable mass and restricted joint mobility may also occur. Systemic symptoms, such as fever and weight loss, are rare. Pathological fractures happen in 10 – 15% of patients. At the time of diagnosis, metastases are present in 15 – 20% of cases, most frequently in the lungs, less frequently in the skeletal system and rarely in lymph nodes. (30,31)

Approximately 80% of tumors are located in the extremities, with the distal femur, proximal tibia, and proximal humerus being the most common sites. Cases in the axial skeleton are rare. (4,30,31)

The 5th edition of the WHO classification system (21) distinguishes between six different subtypes: OS not otherwise specified (OS-NOS), low-grade central OS,

parosteal OS, periosteal OS, high-grade surface OS and secondary OS. OS-NOS is further divided into three categories: conventional OS, which accounts for 93% of all cases, telangiectatic OS, which accounts for 4.5% of cases, and small cell OS, which makes up less that 1% of cases.

Histologically, OS-NOS can be osteoblastic, chondroblastic, or fibroblastic, depending on the produced matrix, and can therefore present with different radiological patterns. (23)

Radiological diagnosis begins with conventional radiography, which can show patterns of mixed osteolytic and osteoblastic lesions. Spiculated periosteal reactions, Codman's triangle and sunburst sign are common periosteal findings. MRI can be used to assess intramedullary extension and soft tissue infiltration, while CT is mainly used for lung staging. The definitive diagnosis is made by biopsy, ideally performed by the same surgeon who will later perform the tumor resection. (31,32)

2.2.2 Ewing sarcoma

Ewing sarcoma (ES) belongs to the *Ewing Sarcoma Family of Tumors (ESFT)*. This family includes classic Ewing sarcoma, Askin tumor (a thoracic manifestation), and peripheral primitive neuroectodermal tumors. These tumors show a t(11;22),(q24;q12) translocation, resulting in an EWS-FLI1 gene fusion in approximately 85% of cases. The remaining ESFTs have a fusion of EWSR1 with another ETS transcription factor, such as ERG, ETV1, or E1AF. (14)

Similar to OS, the initial symptom is usually pain. Localized swelling occurs often, and a palpable mass can be felt. General symptoms such as fever, weight loss, and reduced levels of performance may also be present but are less common. (33)

Around 25% of tumors are primarily located in soft tissue. Approximately 25% of patients have metastases at the time of diagnosis, with pulmonary metastases making up 50%, skeletal metastases 25% and bone marrow metastases up to 20%. (14)

ES shows a broader pattern of distribution than OS. While the extremities are frequently affected, accounting for around 45% of cases, tumors also regularly occur in the thorax, especially on the ribs, and in the pelvis. (4)

Conventional radiography usually reveals osteolytic destruction in the form of 'moth-eaten' or permeative pattern. A characteristic feature is a multilamellar periosteal reaction called 'onion-skin' phenomenon. Spiculated, 'hair-on-end' or Codman triangle periosteal reactions may also be present. ES may not be apparent on conventional radiography when cortical involvement is absent, as imaging finding can be subtle. (33)

2.3 Benign bone tumors

2.3.1 Epidemiology of benign tumors

As established bone tumors are rare overall, but when they occur, they are often benign. Similar to malignant tumors, benign tumors are also found to be more prevalent in males, with the exception of GCTB and ABC. (8) *Table 2* provides an overview of age and gender distribution of benign bone tumors.

	Typical age range in years	**Gender ratio male to female**
Osteogenic		
Osteoid osteoma	5 – 25	2 – 4 : 1
Osteoblastoma	5 – 30	2 – 3 : 1
Chondrogenic		
Osteochondroma	10 – 20	3 : 1
Chondroblastoma	5 – 25	2 : 1
Enchondroma	15 – 35	1 : 1
Chondromyxoid fibroma	10 – 30	1 : 1
Others		
Simple bone cyst	5 – 20	2.5 : 1
Aneurysmal bone cyst	5 – 20	1 : 1.16

Giat cell tumor of bone	20 – 40	1 : 1.2
Non-ossifiying fibroma	5 – 20	2 : 1
Fibrous dysplasia	5 – 10	1 : 1
Hematopoietic		
Langerhans cell histiocytosis	0 - 10	1.5 : 1

Table 2 – Age and gender distribution of benign bone tumors
Adapted according to (8,16,18,34–39)

The most relevant childhood benign bone tumors will be discussed in detail in the following chapters. A particular focus will be placed on X-ray diagnostics.

2.3.2 Osteogenic tumors

2.3.2.1 Osteoid osteoma - OO

Osteoid osteoma (OO) is a benign, yet symptomatically noticeable bone tumor. It typically presents with nocturnal pain that is responsive to NSAIDs. (8,18)

OO account for 10 – 12% of benign bone tumors, making it the second most common lesion after osteochondroma. (15,16) It usually manifests itself between the ages of 5 and 25 years. Men are affected around two to four times more often than women. (16,18,39)

The lesion occurs primarily in the cortex of long tubular bones, particularly in the proximal femur and tibia. However, it can also affect the bones of hands and feet, as well as the spine. (15) In addition to the cortical location, subcortical, intraarticular or subperiosteal locations are also possible. (18)

Radiologically, it is characterized by a central radiolucent nidus with a diameter of less than 2 cm and variable surrounding reactive sclerosis. The nidus is a roundish-oval, osteoid-rich, well-vascularized tissue that histologically corresponds to immature osteoid and fibrous stroma containing osteoblasts, capillaries, and unmyelinated nerve fibers, thus explaining painful presentation. (16,18)

Prominent sclerosis can make visualization of the nidus difficult in some cases, especially for cortically located tumors. Additionally, conventional X-ray sensitivity

may be limited in early stages of the disease or intraarticular locations, therefore requiring CT examination. (8,16,18)

2.3.2.2 Osteoblastoma – OBL

Osteoblastoma (OBL) is a rare, intermediate-grade bone tumor that is locally aggressive. (21,23)

Histologically, it is similar to OO, but it is larger than 2 cm in diameter. It also presents with dull aching pain, but without increased pain at night, and responds less well to analgesics. (18)

Epidemiologically, OBL occur more frequently between the second and third decade of life and it is twice as prevalent in male patients as in female patients. However, the incidence of OBL is around three times lower than that of OO. Another difference is that OBL develops more frequently in the spine and less frequently in the pelvis or extremities. (15,18,39)

Radiographically, OBL exhibit various patterns. Similar to OO, it can appear as a large, oval, radiolucent lesion with irregular margins and surrounding sclerosis. On the other hand, it can present as an expansive lesion with a mineralized matrix and narrow transition zone, especially in the spinal region. Another manifestation occurs in long tubular bones as an aggressive growth that destroys cortical bone and infiltrates soft tissue. For this reason, OBL has also been classified as an intermediate tumor. The size can vary between 2 to 19 cm. Signs similar to those of an aneurysmal bone cyst are seen in approximately 16% of all OBL lesions. (18,39)

2.3.3 Chondrogenic tumors

2.3.3.1 Osteochondroma – OCH

Osteochondroma (OCH) is the most common benign bone tumor, accounting for around 20 – 50% of cases. (35,40) It usually develops in childhood and adolescence and affects male patients thrice as often as female patients. OCH usually only grow during skeletal maturation and show not further growth afterwards. (18,40)

They are often referred to as osseocartilaginous exostoses or simply exostosis, but these terms are no longer recommended by the latest WHO classification. (21)

Up to 85% of patients show solitary OCHs and 15% have multiple OCHs. The multiple form is part of the autosomal-dominantly inherited hereditary multiple exostoses (HME) disease and is associated with a higher complication rate. (35,40) Genetic loss-of-function mutations in the EXT1 and EXT2 genes have been identified, particularly in HME. Due to these genetic alterations, OCH can be assumed to be a true neoplasm and not just a growth disorder or displacement as previously supposed. (35,41)

In most cases, OCH does not cause any symptoms. It can be visible as a painless swelling, which can be a cosmetic concern. Symptoms that are clinically relevant arise when the position or size of OCH affects neighboring structures, causing mechanical irritation, pain during movement, joint misalignment, and compression of nerves and vessels. It can also lead to the formation of a bursa.

A concerning symptom is the sudden growth after epiphyseal joint closure, as this may indicate a malignant transformation into secondary chondrosarcoma. Overall, malignant progression is rare and occurs in about 1% of solitary OCH. In HME-related osteochondromas, the risk increases to 5 – 10%. (18,35,40)

OCH usually develop in the metaphysis of growing tubular bones. The lower extremities, particularly the knee, are frequently affected. They typically grow away from the bone surface in the opposite direction of the nearest joint. (15,18,40)

Diagnosis is primarily based on conventional radiography, in which OCH appear as bony, broad (sessile) or pedicle-based (pedunculated) projection that extends from the medullary cavity and cortex of the original bone. Detecting this continuity in long tubular bones is pathognomonic for OCH. In more complex, flat bones, such as the pelvis, diagnosis can be more difficult and may require additional imaging. An adjacent cartilage cap is present on the process, but it can only be assessed to a limited extent on radiographs, which may make CT or MRI diagnostic necessary. (40)

In children, cartilage caps over 3 cm are considered suspicious for malignancy; in adults this threshold is 2 cm. (35)

2.3.3.2 Chondroblastoma – CBL

Chondroblastoma (CBL) is a rare, mostly benign but locally aggressive bone tumor that occurs predominantly between the ages of 5 and 25. Epidemiologically, CBL account for less than 2% of all BBT. Male patients are affected twice as often as female patients. (18)

CBL is typically centrally located in the epiphyses of the long tubular bones, particularly the proximal femur, distal femur, proximal humerus and proximal tibia. Less frequently, it may affect the short bones of the hands and feet or the vertebrae. Clinically, CBL usually presents as pain, swelling, and restricted movement in the affected joint, due to its proximity to the joint. (18)

Conventional radiography shows well-defined lytic lesions accompanied by a narrow zone of sclerosis. According to the literature, sign of an aneurysmal bone cyst-like lesion can be detected in 15 – 30% of cases. (15,18)

2.3.3.3 Enchondroma - ENC

Enchondroma (ENC) is another common tumor in the general population. It develops during childhood and adolescence, but the peak incidence is the third decade of life, when the lesions are often discovered incidentally. (18)

Enchondromatosis is defined as the presence of multiple enchondromas, as seen in Ollier disease or Maffucci syndrome. The overall potential for malignant transformation from ENC to chondrosarcoma is low, but it is increased to up to 50% in the case of enchondromatosis. Genetically, IDH1 and IDH2 gene mutations are frequently reported in ENC and enchondromatosis. (16,42)

Enchondromas are made up of hyaline cartilage. They are located medullary in the small tubular bones of the hands and feet, and occasionally in the long tubular bones. Radiologically, they present as well-defined, round, radiolucent foci with a thin cortex. They may also exhibit chondral calcifications that appear as ‘rings-and-arcs’. This mineralization of the matrix may be absent, particularly in younger

patients. A CT scan can be used to more accurately assess calcification patterns. (16,18)

2.3.3.4 Chondromyxoid fibroma – CMF

Chondromyxoid fibroma (CMF) is a rare tumor that accounts for less than 1% of all bone tumors. (18) Formerly classified as an intermediate (locally aggressive) tumor, CMF has been downgraded to a benign lesion since the 5th edition of the WHO classification. (21,23)

The tumor is most often located eccentrically in the metaphysis of long tubular bones, particularly around the knee in the proximal tibia, distal femur, and proximal fibular. The foot and pelvic bones may also be affected. (18)

Radiographs typically show an eccentric, lytic lesion in the metaphysis with irregular and often sclerotic margins. Cortical thinning or perforation is often visible, periosteal reaction is rare. (18)

2.3.4 Other bone tumors

2.3.4.1 Simple bone cyst – SBC

The simple bone cyst (SBC), also known as unicameral, solitary, or juvenile bone cyst, is a benign, fluid-filled bone lesion that predominantly occurs in growing children and adolescents. Contrary to its synonymous names, the simple bone cyst can also be multilobulated, rarely multiple and sometimes also occur in adults. (18)

SBC most frequently develop during the first two decades of life. Male adolescents are affected more often than females, at a ratio of 2.5:1. Clinically, SBCs are often asymptomatic and are discovered incidentally or following a pathological fracture. (18)

The most common locations are the metaphyses of the long tubular bones, particularly the proximal humerus (50% of cases) and the proximal femur (25% of cases). Rarely, the pelvic or tarsal bones are affected. (8,15,18)

Radiographically, SBCs appears as a round or oval-shaped, well-defined lytic lesion surrounded by intact cortical bone. Its central location in the metaphyseal

region is characteristic. In the context of a pathological fracture, the ‘fallen fragment sign’ may be present, showing an intracystic detached bone fragment. (15,18)

2.3.4.2 Aneurysmal bone cyst - ABC

The aneurysmal bone cyst (ABC) is an expansive, blood-filled, septated lesion that develops either primarily or secondarily from other lesions. They usually develop around the second decade of life and are very slightly more common in female patients. Although they are classified as benign, they can be locally aggressive. (15,18,36)

Primary ABCs are considered true neoplasms and exhibit a genetic translocation that results in the overexpression of the USP6 gene in most cases, which cannot be found in secondary ABCs. Secondary ABCs develop as a reactive process in the context of other bony lesions such as GCTB, CBL, OBL or FD. (15,43)

Clinically, ABCs typically present with pain that is often load-dependent and may worsen at night and during the course of the disease. In some cases, there is visible swelling or a local increase in size of the affected bone. In rare cases, such as when the spine is involved, neurological deficits like paresthesia or motor deficits is possible. Pathological fractures may be the first clinical sign, especially in larger lesions. (8)

The most common locations are the metaphysis of long tubular bones, particularly the femur, tibia, and humerus. However, the pelvis and spine can also be affected. (15,18)

In comparison to other lesions, which are primarily named according to their histological structure, ABC is named after its radiographical appearance. Radiographs typically show a metaphyseal, eccentric, osteolytic, expansively growing lesion with ballooned (aneurysmatic-expanded) cortical bone. If the lesion is located in the diaphysis, it is more likely to be cortical or subperiosteal. (18) There are multiple phases of growth starting with initial focal osteolysis. During the following active phase, the lesion grows a causes bone destruction, remodeling, thinning of the wall, and a periosteal reaction. Finally, there is a stabilization phase with increasing calcification and ossification of the margin. (44)

If an MRI is performed, it will typically show fluid-fluid levels, that are not specific but highly suggestive of ABC. (45) In addition to MRI, a biopsy is usually necessary for confirmation, especially since it is important to differentiate ABC from telangiectatic osteosarcoma. (46)

2.3.4.3 Giant cell tumor of bone – GCTB

Giant cell tumor of bone (GCTB) is an uncommon, intermediate-grade (locally aggressive and rarely metastasizing) bone tumor that primarily affects young adults in the third decade of life. Women are affected slightly more frequently. (23,37)

The tumor predominantly manifests in the epiphyses of long tubular bones, particularly the distal femur, proximal tibia, and distal radius. (37)

Radiographically, an eccentric lytic lesion with a blurred margin is seen. The epiphyseal lesion may extend into the adjacent metaphyseal area. It is characterized by an absence of a sclerotic border and frequent extension to the subchondral cortical bone. It is also frequently associated with secondary ABCs. (18,37)

2.3.4.4 Non-ossifying fibroma – NOF

Non-ossifying fibroma (NOF) was first described by *Phemister* (47) in 1929 under the term ‘chronic fibrous osteomyelitis’. In 1941, *Sontag and Pyle* (48) provided a comprehensive radiological description of the lesions as ‘cyst-like areas in the distal femur of children’. *Jaffé and Lichtenstein* (49) coined the term ‘non-osteogenic fibroma’ in 1942, which later developed into non-ossifying fibroma, to describe this condition as an independent entity.

Many terms have historically been used to describe NOFs, including histiocytic fibrous defect, metaphyseal fibrous defect, fibrous cortical defect, fibrous xanthoma, benign fibrous histiocytoma, cortical desmoid, and histiocytic xanthogranuloma. (50)

Currently, the terms ‘fibrous cortical defect (FCD)’ and ‘NOF’ are commonly used. There are no histological differences between the two conditions, which both consist of fibrous stroma, giant cells, and xanthomatous histiocytes. They are

distinguished by size and locations. FCDs are smaller and located in the cortex. NOFs are characterized as a lobulated, eccentric lesion with a thinned or partially absent cortex. According to the latest WHO classification, this distinction is no longer recommended and only the term NOF should be used. (18,21)

NOF is the most common benign lesion in children and occurs in between 30 – 40% of all children. They have a high potential for regression and self-healing through remodeling and ossification, thus rarely occurring after the age of 20 years. (15,48)

In most cases, NOFs are an incidental finding that is completely asymptomatic. Asymptomatic lesions should not be treated and are considered a ‘leave me alone’ or ‘do not touch’ lesions as long as they are biomechanically stable. In cases of particularly large lesions that exceed 50% of the transverse diameter, surgical stabilization should be considered. (15,17,51)

NOFs are usually localized in the metaphyseal regions of the long tubular bones, with 55% of lesions found around the knee. The distal femur, proximal tibia, and distal tibia are affected in descending order of frequency. (15,18)

Radiologically, an eccentric, polylobulated lesion with a sclerotic rim and scalloped borders is seen. (8) The diagnosis can be made with almost 100% certainty from radiographical appearance alone. (17)

2.3.4.5 Fibrous dysplasia – FD

Fibrous dysplasia (FD) is a sporadic, non-hereditary bone formation disorder characterized by the transformation of normal bone and bone marrow into fibrous connective tissue with immature bone tissue. It is often an incidental finding. (18)

FD is monostotic in 70 – 80% of cases and polyostotic in up to 30% of cases. The polyostotic form is occasionally associated with syndromic diseases, such as McCune-Albright syndrome (characterized by endocrine disorders and café-au-lait spots) and Mazabraud syndrome (characterized by muscular myxomas). (15,18)

An activating GNAS mutation can be identified, which can be helpful in differentiating FD from low-grade osteosarcoma. (15)

The clinical presentation varies depending on the location and extent of lesions. Most cases are asymptomatic, though pain and pathological fractures can occur. The most common sites of localization include the craniofacial skeleton, femur, tibia, and ribs. (15,18)

Radiographs typically show sharply defined lytic lesions with an internal relief the appears like ‘ground-glass’ with a sclerotic border. The cortical bone structure is often expanded and displaced but rarely penetrated. More severe deformities, such as the ‘shepherd’s crook’ deformation of the proximal femur, are rarely observed. (8,18)

2.3.5 Hematolymphoid disorders – Langerhans cell histiocytosis

Langerhans cell histiocytosis (LCH) is a rare heterogenous disease defined as a neoplasia of myeloid dendritic cells. These proliferating cells infiltrate various organs and tissues, resulting in a wide range of symptoms. (15)

LCH covers a spectrum of diseases. Different entities can be differentiated depending on the location, number of manifestations, and accompanying phenomena. These include eosinophilic granuloma (90% single bone involvement, 10% multiple bones – the most benign form), Hand-Schüller-Christian disease (bone lesions, exophthalmos, and diabetes insipidus – the intermediate form) and Letterer-Siwe disease (skin, visceral and brain lesions – the most severe form). (8)

Radiographically, the eosinophilic granuloma appears as a ‘punched-out’, osteolytic bone lesion without a surrounding bone reaction. It most commonly affects the skull, femur, vertebrae, pelvis, ribs, and mandible. (8)

3. Diagnostic imaging of bone lesions

As part of the diagnostic work-up of bone lesions, a systematic analysis is required to differentiate between benign and malignant neoplasms, as well as tumor-like lesions. The previous chapters described the radiologic-morphologic characteristics of individual entities; the following chapters relate them to each other.

3.1 Diagnostic steps

The diagnostic workup of osseous lesions primarily begins with conventional radiography, as they can provide clear diagnostic clues in up to 80 – 90% of cases. (52)

Bone lesions are often discovered incidentally on radiographs taken for other reasons or, less frequently, as part of cross-sectional imaging. Even when a cross-sectional image, such as MRI, is available, an X-ray examination should still be performed as well because MRI is not specific to most conditions, and the potential malignancy of a lesion is often overestimated. (53)

In case of anatomically complex bone structures, especially the spine and pelvic bones, a CT scan should be performed in addition to conventional radiography. (54,55) However, ALARA principles (as low as reasonably achievable) should always be applied, even more so in the pediatric population, to avoid unnecessary radiation exposure. (56,57)

In cases of suspected malignancy, an MRI examination should follow the initial imaging for treatment planning. For staging, it is recommended that a CT scan of the lungs using maximum intensity projection (MIP) technique should be performed to detect lung metastases. A 18F-FDG PET/CT (18F-fluorodexoglucose positron emission tomography/computed tomography) scan is recommended for the detection of extra-pulmonal, in particular bone, metastases. Recent studies have shown the superiority of 18-FDG PET/CT over Tc-99m bone scintigraphy in identifying extrapulmonary metastases. (58)

After staging, a biopsy should be performed to make a final diagnosis. (58,59)

3.2 Radiological features of benign and malignant lesions

Several guiding questions have been established to assess bone lesions. These questions should be considered and answered systematically. (8,39,55)

- In which bone is the lesion located?
- Where in the bone is the lesion located?
- What is the lesion doing to be bone?

- How does the bone react to the lesion?
- Is there a periosteal reaction?
- Is there a soft tissue component?

Clinical features such as age, gender, ethnicity, and symptoms also contribute to the differential diagnosis. In particular, age is the most important factor in differential diagnosis, as many lesions exhibit age-specific distributions, as shown in *Table 2* in chapter *2.3.1 Epidemiology of benign tumors*.

3.2.1 Location

The location of an osseous lesion is a fundamental diagnostic clue. In addition to identifying the affected bone, the location of the lesion within in bone is crucial. In long tubular bones, the epiphysis, metaphysis, and diaphysis are distinguished along the longitudinal axis, and the cortical, medullary (central), and eccentric location are distinguished along the short axis (*Figure 1*)

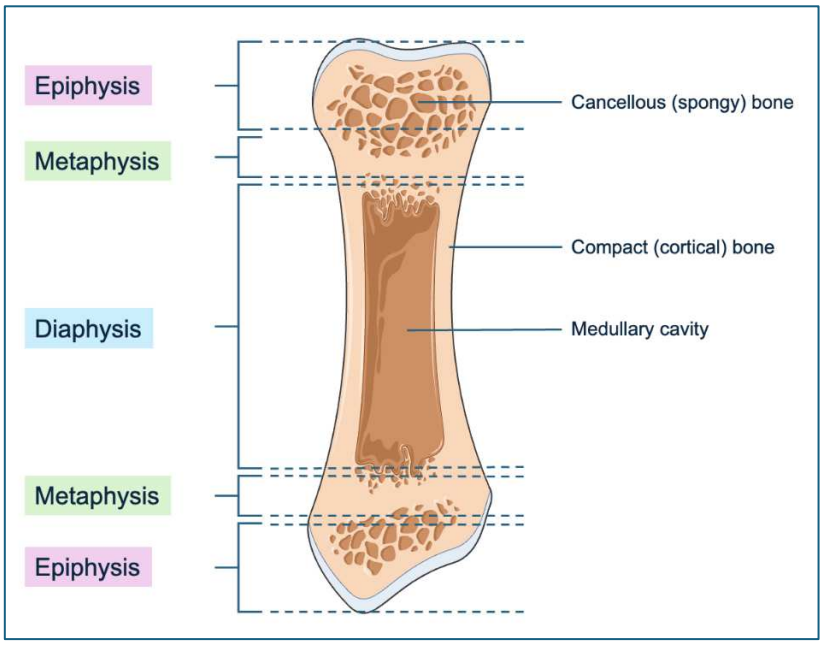

Figure 1 – Sections of a bone
Image adapted from Servier Medical Art (https://smart.servier.com/), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Table 3 shows the most common locations of bone tumors in relation to the affected bone, long axis, and short axis.

Matrix	Tumor	Location in Skeleton		Location within Axis	
		Most common	*Other*	*Long*	*Short*
Osteoid	OO	Femur, tibia	Spine, hand, foot	Diaphysis	Cortex or adjacent
	OBL	Spine	Long bones, pelvis, hands, feet	Diaphysis	Central
				Metaphysis	
	OS	Around knee, humerus	Pelvis, Spine, Ribs	Metaphysis	Depending on subtype
Chondroid	OCH	Around knee	Humerus, pelvis, scapula	Metaphysis	Surface
	CBL	Femur, tibia, humerus	Hands, feet, spine	Epiphysis	Eccentric
	ENC	Hands and feet	Long bones	Metaphysis	Central
	CMF	Around knee	Foot, pelvis	Metaphysis	Eccentric in long bones, central in short bones
Other	SBC	Humerus, femur	Tibia, calcaneus	Metaphysis	Central
	ABC	Femur, tibia, humerus	Spine, pelvis	Metaphysis	Eccentric
				Diaphysis	Cortex or adjacent
	GCTB	Femur, tibia, radius	Humerus	Epiphysis	Eccentric
	NOF	Around knee	Humerus, ulna	Metaphysis	Eccentric
	FD	Skull, facial bones femur, tibia	Ribs	Diaphysis	Central

Hematopoietic	LCH	Skull, femur, spine	Pelvis, ribs, mandibula	Diaphysis	Central
Small-round-cell	ES	Femur, pelvis, humerus	Chest wall, ribs, spine	Diaphysis	Central

Table 3 – Localizations of primary bone tumors by tumor site, long axis and short axis
Modified after (18) and adapted according to (10,15,39,40,55)
OO – Osteoid osteoma, OBL – Osteoblastoma, OS – Osteosarcoma, OCH – Osteochondroma, CBL – Chondroblastoma, ENC – Enchondroma, CMF – Chondromyxoid fibroma, SBC – Simple bone cyst, ABC – Aneurysmal bone cyst, GCTB – Giant cell tumor of bone, NOF – Non-ossifying fibroma, FD – Fibrous dysplasia, LCH – Langerhans Cell histiocytosis, ES – Ewing sarcoma

The development of bone lesions depends on constant growth in the remodeling zones, which are predilections for neoplasia. This is illustrated by the fact that the list of metaphyseal differential diagnoses is significantly longer than the lists of epiphyseal and diaphyseal diagnoses, as most remodeling takes place in the metaphysis. (10)

3.2.2 Margins and growth patterns

The margins of a lesion and the transitional zone are particularly important for diagnosis and assessing biological aggressiveness. A lesion with sharply defined margins and narrow transition zone usually indicates a benign, non-aggressive process. In particular, a sclerotic border is a sign of slow growth. (1,10)

Two different bone reaction patterns are described: osteolysis and new bone formation. According to *Lodwick* (60), osteolysis can be divided into three types of destruction: geographic, moth-eaten, and permeative.

Type I, the geographic type, describes focal, discrete, roundish to oval-shaped areas of complete bone destruction with well-defined borders. Based on the margins, type I can be further differentiated into three subgroups. Type IA corresponds to the least aggressive findings and appears as a lesion with a sclerotic rim. Examples of this subgroup include NOF, FD, and CMF. Type IB also shows well-defined tumors but lacks the sclerotic rim. This form can occur in both benign and malignant lesions. Examples include GCTB, ABC, multiple myeloma

and low-grade chondrosarcoma. Type IC shows a blurred boundary with a wide transition zone and is usually indicating a locally aggressive or malignant process. Classic examples are chondrosarcoma and osteosarcoma. (1,60)

Type II, the moth-eaten type, appears as multiple, blurred osteolytic regions with jagged edges. Type III is characterized by permeative bone destruction, which manifests as diffuse, worm-like destruction and infiltration of the medullary cavity. (1,60) Both types often indicate aggressive growth but can also occur in nonmalignant diseases, such as osteomyelitis or eosinophilic granuloma. (1,54)

3.2.3 Matrix mineralization

Many tumors are named after their matrix formation, which is derived from histology. Imaging can also provide indications of the underlying histology. (54) An osteoid-producing tumor shows an ordered trabecular network. Chondrogenic tumors show punctiform, flaky calcifications that may appear in a ‘ring-and-arc’ pattern. Fibrous tumors often show a ‘ground-glass’ structure, which describes hazy, but mostly homogenous radiodensity. An increase in mineralization over time may indicate that the lesion is maturing and is most likely benign (1)

In terms of matrix, conventional radiography has limitations. Approximately 40 – 50% of the trabecular bone mass must be lost before a lesion becomes visible. Therefore, CT and MRI scans are useful for better assessing tumor margins and matrix structure, but also to evaluate periosteal response and soft tissue extension. (1) ALARA criteria for CT examinations and the need for sedation in younger children for MRI examinations should be considered.

3.2.4 Periosteal reaction

The periosteal response is a direct consequence of cortical involvement, vascular stimuli, or an inflammatory response. There are three patterns: continuous, interrupted and complex.

The continuous periosteal reaction shows a smooth reaction along the bone and can be further divided into solid/ unilamellar form, or multilamellar (‘onion-skin’) form. Interrupted periosteal reactions occur with cortical perforation. Spiculae and

‘hair-on-end’ or ‘sunburst’-like configurations may be present. A Codman triangle describes a partial elevation of the periosteum during rapid growth. (1,39,54)

3.2.5 Soft tissue involvement

The presence of soft tissue involvement strongly indicates potential malignant genesis. However, exceptions exist, such as benign lesion like GCTB and inflammatory processes like osteomyelitis, which can also have a soft tissue component. (1)

Soft tissue is better judged in MRI, but as mentioned before it must be considered that sedation might be necessary in younger children.

3.3 Diagnostic pitfalls and differential diagnosis

As shown in *Table 4*, certain growth patterns often allow to estimate a lesions’ benignity or malignancy. However, these rules are not universal. Radiological features of different entities often overlap, making it difficult to distinguish between benign, intermediate, and malignant tumors. This is especially challenging in a pediatric collective due to bone maturation processes, variable growth zones, and anatomical variants.

Feature	**Usually nonaggressive**	**Usually aggressive/ undetermined**
Type of osteolysis:	Type I: Geographic	Type II/ III: moth-eaten/ permeative
Margins:	Well defined	Blurred
Transition zone:	Narrow	Wide
Periosteal reaction:	Continuous, unilamellar	Multilamellar, interrupted (sunburst, hair-on-end, Codman)

Soft tissue involvement:	Absent	Present

Table 4 – Typical radiological features of nonaggressive and aggressive lesions

Diagnostic pitfalls occur when lesions that initially appear harmless turn out to be malignant. For example, low-grade osteosarcomas have an unremarkable radiological appearance. (61) The opposite can also occur, when benign changes appear aggressive and are mistaken as malignant lesions. In particular, ABCs can be very similar to telangiectatic osteosarcoma on radiological imaging (lytic, expansive lesion, thin cortex, septated internal structure). (62)

Inflammatory of infectious processes, such as acute osteomyelitis, can mimic permeative destruction in early stages. (63) In such cases, anamnesis, laboratory parameters, and follow-up examinations are essential for reliable differentiation.

Certain bone regions, such as the pelvis, spine, and shoulder girdle, present addition diagnostic challenges. Due to their complex anatomy and overlapping of structures, lesions can be easily overlooks or misinterpreted. (64)

Physiologically, age-related changes can also lead to misinterpretations, as normal variants can be mistaken for pathology. (53)

It becomes apparent that the diagnostic evaluation of primary bone lesions is very complex, requiring expertise and experience. Even experienced radiologists can reach their diagnostic limitations in challenging cases. The use of artificial intelligence, particularly deep learning-based methods provide a promising approach, as automated image analysis and pattern recognition can detect subtle features that are challenging for human observers. The following chapter is therefore dedicated to the fundamentals and currently researched applications of AI in medical imaging, with a focus on deep learning in musculoskeletal oncology.

4. Artificial intelligence in medical imaging

4.1 General concepts of AI

Artificial intelligence (AI) is an interdisciplinary field of research that deals with the development of computer systems with capabilities typically associated with

human or animal intelligence. These include perception, reasoning, problem solving, language comprehension, decision-making, and creative processes. A key advantage of AI-based systems is their ability to efficiently analyze and process large and complex amounts of data. (65)

The history of AI in medicine stretches back over 50 years. One of the first expert systems developed to support the diagnosis and treatment of bacterial infections was MYCIN, which was created in the 1970s. (66) Today, the range of medical AI applications extends from radiological image analysis (67,68)**,** personalized therapy recommendations (69–72)) to population-based prediction models in public health (73,74). Of particular relevance to this thesis is the work in the field of radiological image analysis in relation to bone tumors, which is discussed in *Chapter 4.2 Recent studies on AI-assisted diagnostics in bone tumor imaging*.

A central subfield of AI is machine learning (ML), which in turn represents the currently most relevant area of AI research, deep learning (DL). Convolutional neural networks (CNNs) are special DL-models that are used for image recognition in particular. Computer vision (CV) is one application area of CNNs, which is suited for tasks such as image classification, object recognition, and segmentation of image areas. (65,75,76) The diagram in *Figure 2* shows the hierarchical relationship between AI, ML, DL, and CNNs, as well as CV as one form of application of CNNs.

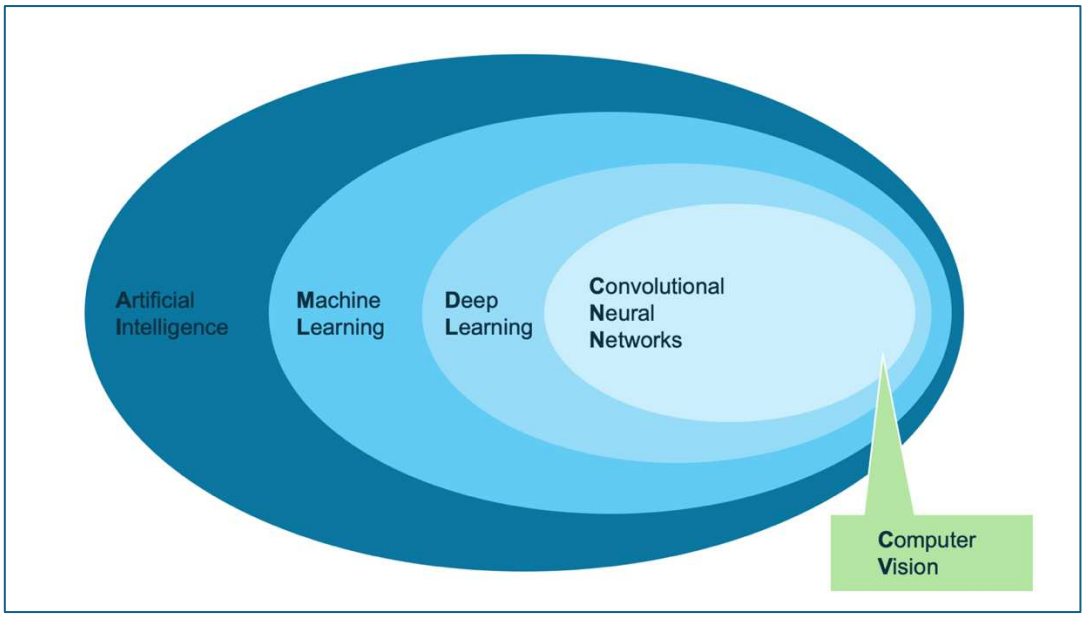

Figure 2 – Architecture of the hierarchical relationship and Computer Vision being an application of CNN Modified after (77)

ML is the process by which computers are able to recognize patterns, model correlations, and make predictions by accessing large amounts of data, without every possible decision having been explicitly programmed. (65,78)

There are three main types of learning in ML: supervised, unsupervised and reinforcement learning. In supervised learning, the model is trained with labeled data, where both input and target variables are known. Typical applications include classification and regression, for instance in medical image analysis or in laboratory diagnostics. In unsupervised learning, no target variables are specified, and thus the focus is on discovering structures or patterns in the data, for example through clustering or dimension reduction. Reinforcement learning involves interaction with an environment in which the model adapts its strategy through feedback in the form of rewards or punishments. (65,78)

DL is a specialized form of ML. It is based on artificial neural networks whose architecture is in turn based on the information processing of biological nervous systems. DL models consist of several hierarchically layered levels of artificial neurons. In contrast to traditional ML methods, where relevant features have to be defined manually in advance, DL models are able to learn such features directly from the raw data. (78) A neural network typically consists of an input layer, a variable number of hidden layers, and an output layer. Networks with only one

hidden layer are referred to as ‘shallow’, while those with several layers are referred to as ‘deep’. The performance of such models improves with access to large, high-quality datasets. (76,78)

A significant technical advance was the development of CNN by *LeCun et al.* (79) in 1989. CNNs are particularly well-suited to processing visual data, using convolutional layers to recognize local features, such as edges, textures, and geometric shapes. In deeper layers, these features are combined to form more complex, abstract representations. Pooling layers reduce the spatial resolution of the feature maps, thereby reducing the computational load and improving the models’ generalizations capability. (78)

Building robust models requires careful data set structuring. The standard approach involves dividing the data into non-overlapping training, validation, and test sets. (80)

- ‘Training set’: The largest subset of images, which is used for model training
- ‘Validation set’: This subset is used to validate learning progress during training phase
- ‘Test set’: This set includes images not encountered during the training or validation procedure. Ideally, this set is supplied by an external institution to verify the model’s generalization capabilities

Distortions in the data distribution, for instance due to unequal distribution of sample groups or visible characteristics, can significantly impair the model’s performance and should be considered when curating the test set. (80)

A key advantage of algorithmic systems is their robustness against factors such as fatigue, attention fluctuations, and subjective experience. These are particularly relevant given the global shortage of radiologists. However, technological challenges remain, including limited generalizability, unpredictable model behavior, and insufficient training data. (81,82)

A central problem in the application of ML and DL methods is overfitting, in which the model adapts too strongly to the specific characteristic of the training data,

which makes generalization to new, unknown data more difficult. This can be caused by a training dataset that is too small, an excessive number of input variables (high dimensionality) or an overly complex model architecture. (65,83) This can be counteracted though regularization, drop-out, cross-validation or by using extensive and representative datasets. (80,83)

Data augmentation and transfer learning are two methods used to increase the performance of neural networks. Data augmentation refers to targeted image transformations, such as rotating, scaling, mirroring, shifting, or brightness adjustments and is used to increase the variety of the training data without generating new image information. (83) Similarly, generative adversarial networks (GANs), which generate realistic-looking synthetic image data to expand the training set, can be used. (84) Transfer learning uses pre-trained networks whose weights were learned from large, general image datasets (for example ImageNet) and then fine-tuned for a specific target application. This saves computing time and resources, while increasing the model's performance, especially with smaller datasets. (76)

The explainability of AI models is one of the key requirements for clinical use and regulatory approvals. A major obstacle here is the so-called 'black box' problem. This refers to the fact that many DL models, especially CNNs, make decisions based on highly complex internal processes, that are not immediately comprehensible to human users. This hinders the identification of sources of error, the assessment of risks, and the establishment of trust in AI systems among specialists. In medical diagnostics, transparency is essential in order to gain the trust of users and make the model's decisions comprehensible. Various explainable AI (XAI) techniques have been developed to address this lack of transparency. (85) Examples include Deconvnet (86), Guided Backpropagation (87), Class Activation Mapping (CAM) (88), which has been further developed into Grad-CAM (89) and Grad-CAM++ (90). In CAM, activation patterns in the final convolutional layer are linked to the weight of the classification layer to show which image regions contribute to the decision for particular class. The result is displayed as a heatmap, highlighting the relevant areas in color. (88) In the following development, Gradient-weighted CAM (Grad-CAM), the direction and amount in

which the model parameters adjusted during training, the so-call gradients, are used to determine the significance of individual feature maps for a decision.(89) Grad-CAM++ further refines this approach by taking not only individual gradients, but also the weighting of multiple positive gradients into account. This allows for the creation of more relevant and stable heatmaps, which visualize whether the model is actually using disease-relevant structures for its prediction. (90)

4.2 Recent studies on AI-assisted diagnostics in bone tumor imaging

In recent years, AI has established itself as a promising tool in the field of musculoskeletal radiology. Research in the specialized field of imaging is primarily focused on bone age determination and fracture detection (77,91), but progress is also being made in oncological imaging, like the diagnosis and classification of PBTs (92–94). Of the 272 CE-certified (Conformité Européenne) AI tools currently available, 42 are approved for use in the musculoskeletal field. Of these, only five are explicitly approved for pediatric use, four for bone age determination and one for fracture detection. (95) Research implementing AI in the field of pediatric radiology lags significantly behind the strides in general radiology. The reasons for this include limited data availability, economic factors, and significantly different clinical characteristics of pediatric patients, which makes transferring ‘adult’ AI tools difficult. (96)

This chapter provides an overview of recent studies on the use of DL to evaluate X-ray images and differentiate between benign and malignant bone tumors. A structured PubMed search shown in *Figure 3* identified seven original papers on this topic (97–103). One paper was excluded because the full text could not be found (102). Another was excluded for solely focusing on the detection, but not the classification on tumors (103).

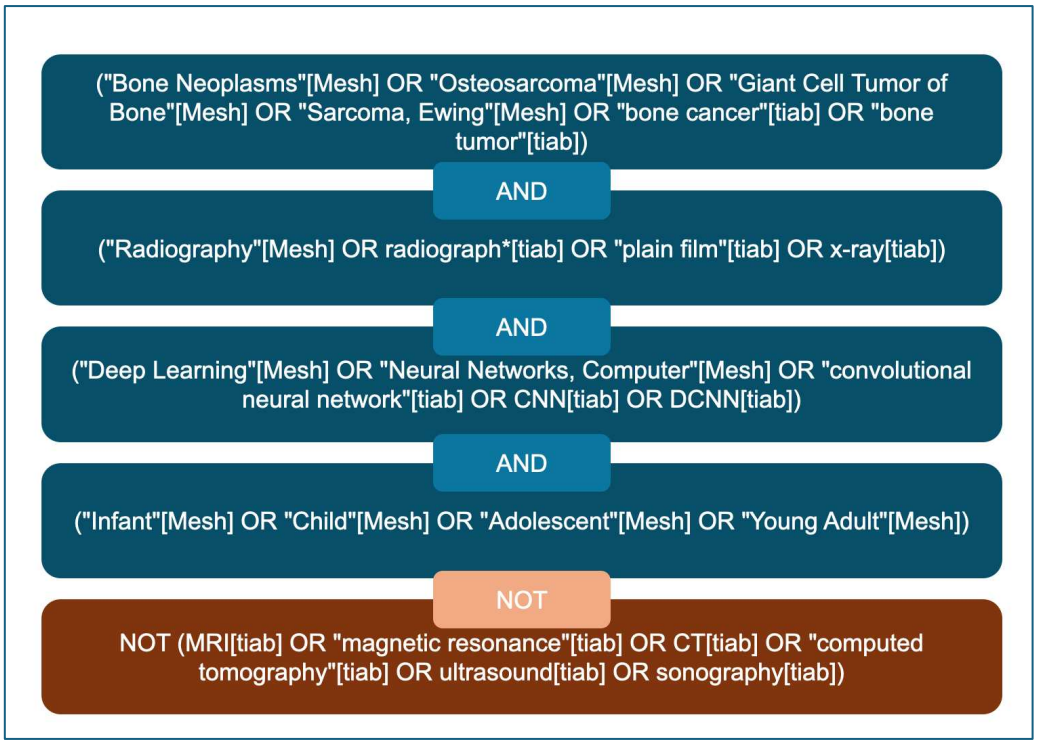

Figure 3 – PubMed search term

Additional free-text searches were conducted, using combinations of the aforementioned terms and additional terms like ‘musculoskeletal’, ‘oncology’, ‘radiology’, ‘pediatric’ and ‘artificial intelligence’ as well as backward citation searching, yielding 8 additional studies (104–111).

This literature search focused specifically on papers dealing with the use of DL-applications, radiographs, and, where possible, ought to include children and/or adolescents and is not meant to be exhaustive. Studies on other modalities such as CT, MRI, or PET-CT exist, but were excluded for the sake of clarity. Furthermore, purely technical studies without clinical relevancy were deliberately excluded in order to focus on practical applications in everyday pediatric radiology.

A Classic CNN, namely EfficientNet B0, was examined in the study by *He et al*. (97) in 2020 for their three-way classification results of PBT-images into benign, intermediate, and malignant. The results were compared to those of radiologists of varying levels of experience, showing that purely CNN-driven classification is comparable to that of experienced radiologists and superior to that of junior radiologists.

A study be *Liu et al.* (98) in 2022 compared the three-way classification of a DL approach and a DL-ML-fusion approach that included clinical data. They also compared both approaches with the performance of five radiologists and the performance of the radiologists with assistance of the fusion-model. The DL model proved inferior to the DL-ML fusion approach. In addition, there was a potential benefit for the diagnosis of the five radiologists assisted by the fusion model.

A similar study by *von Schacky et al.* (106) in 2022 also compared the performances of a multilayer ANN (artificial neural network) combined with demographic features, radiomic features or both, and showed that the combination of both with the ANN improved accuracy, sensitivity, and specificity.

All three studies dealt with fairly large cohorts (He: 2899 images, Liu: 982 images, von Schacky 880 patients), which however included adult patients in addition to pediatric patients. *He et al*. performed an age-stratified calculation, which tended to show better results in patients under the age of 36, potentially due to the fact that the majority of training data consisted of pediatric patients. *Liu et al*. found significant differences between dignity subgroups. An age-stratified calculation was not performed. *Von Schacky et al*. provided no age distribution across different entities.

In 2024, *Hinterwimmer et al*. (101) publish a study on the connection between DL and clinical metadata. The study showed that the model improved with reference to metadata, similar to the aforementioned studies (98,106).

Hinterwimmer et al. (100) also presented a study about a recommender-based approach to classify PBTs in the same year. The model recommends a diagnosis based on similarities to previous correctly classified patient cases.

Many studies focused on a classification based on biological characteristics (benign, intermediate or malignant). An attempt of a multi-class classification into entities was made by *von Schacky et al.* (105) in 2021. They used a pretrained multitask DL model, which worked well for binary classification (malignant/ non-malignant), but left room for improvement for the classification into 16 different entities with an accuracy rate of 42.9%. The authors also compared the performance to radiologists, who performed at 62.2% accuracy for

musculoskeletal fellowship-trained radiologists (n = 2), and 44.1% accuracy for radiological residents (n = 2) for the classification into distinct entities.

Most studies included images from various regions of the body, but some were limited to specific body parts. For example, *Guo et al.* (110) published a study in 2024 focusing on classification on malignant PBTs of the spine. As another example, a study by *Do et al.* (104) in 2021 focused on the detection and classification of PBT around the knee. A different study by *Park et al*. (107) in 2022 focused on the proximal femur. Both studied aimed for a three-way classification into healthy, benign and malignant subgroups. A four-way classification was performed by *Li et al*. (109) in 2023 by adding the intermediate subgroup, not focusing on a specific part of the body.

All of the studies mentioned so far included either both adults and children or no information on the demographics of the study population was given. Interpreting X-ray images of children is not the same as interpreting images of adults, since children's bones are still growing and constantly changing. Studies with an exclusively pediatric population were rare and only two papers matching the search profile were found.

In 2022 *Consalvo et al.* (99) published a study focusing on the differentiation between Ewing sarcoma and osteomyelitis in pediatric patients. The study was structured into two phases. In the first phase, cases were classified as 'healthy' or 'pathological'. The accuracy, sensitivity, and specificity rates were 90.6%, 89.4%, and 91.0% respectively on external testing. In the second phase, cases were classified as either Ewing sarcoma or acute osteomyelitis. For this phase, the external testing rates were 86.7%, 100.0%, and 76%. It should be noted that the study population of 58 patient, 31 of which were healthy controls, was rather small.

Breden et al. (108) conducted another study involving only pediatric patients in 2023. The authors used a pretrained Vision Transformer model to investigate the early detection of childhood tumors around the knee.

Another study that focused on the differentiation of PBT and bone infection was performed recently, in 2025, by *Wang et al*. (111). The authors compared DL-models and a DL-ML-Ensemble Model but included again both adult and pediatric

patients. There was a statistically significant difference between the ages of the tumor and the infection group. Age-stratified calculations were not performed.

Many studies mentioned limited data as well as class imbalance as primary limitations and used pretrained models and data augmentation to combat those obstacles.

A study that solely focuses on the classification of PBT in pediatric bones could not be found.

5. Materials and methods

Ethical approval for the study (No. EK 1178/2024) was obtained from the Ethics Committee of the Medical University of Graz. Informed consent was waived due to the retrospective design of this study. All methods were carried out in accordance with the Declaration of Helsinki and the relevant regulations and guidelines. The dataset may be made available upon reasonable request.

5.1 Population, in- and exclusion criteria

The radiographs used for this thesis were taken at the Division of Pediatric Radiology of the Department of Radiology at the Medical University of Graz, Austria. Images which were taken between 01.01.2004 and 30.05.2024 were included. Patients were included if they were between 0 - 19 years old at the time of examination and had a coded diagnosis of a primary benign, intermediate or malignant bone lesion. The diagnoses were either confirmed by typical radiological finding or, in cases where this is relevant, by a histopathological diagnosis.

Only the imaging findings at the initial presentation of each patient were included. Therefore, imaging after biopsies, interventions, chemo- or radiotherapy were excluded. Patients were also excluded if no bone lesion could be confirmed after the initial suspected diagnosis.

The data set contained 800 images (707 [88.37%] of benign tumors, 93 [11.62%] of malignant tumors) of 228 individual patients (197 patients with benign tumors, 31 patients with malignant tumors).

Patients age ranged from 0.10 to 17.60 years, with a mean of 8.41 years and a standard deviation of 4.26 years. The mean age of patients with benign tumors was 8.23 years, that of patients with malignant tumors 9.77 years.

In the overall cohort, 38.25% of the patients were female (n = 396) and 61.75% were male (n = 494). A similar sex distribution was observed among patients with benign tumors, with 39.32% (n = 278) being female and 60.68% (n = 429) male. In contrast, the proportion of female patients was lower at 30.11% (n = 28) in malignant cases, whereas male patients accounted for 69.89% (n = 65) of cases.

5.2 Image processing and data anonymization

The images used were acquired at the Division of Pediatric Radiology of the Medical University of Graz and stored in the local PACS (Picture Archiving and Communication System) as DICOM (Digital Imaging and Communications in Medicine Format) files. After selecting the images, the corresponding DICOM files were exported and converted to PNG files (16-bit, grayscale).

To deidentify the data, each image was given a new file name. These were formed from the following metadata values:

- SHA-3-256 hash of patient ID
- Unix-timestamp of examination time minus 8-digit integer hash of the patient ID
- Study number
- Region code
- Sex
- Age in years (rounded to 1 decimal place)

The Secure Hash Algorithm-3 (SHA-3) describes a procedure in which an arbitrarily long input file (message) is output as a value (hash value or digest) of fixed length. In our case, SHA-3-256 was used, and the digest output corresponds to 256-bit or rather 64-hex characters. (112) The advantage of hashing is that the output data cannot be converted back to the original data.

The Unix timestamp is a way of representing a particular point in time as the number of seconds that have elapsed since 1 January 1970 at 00:00:00 UTC.

UTC refers to Coordinated Universal Time, which means this system works independently of time zones. For further deidentification, an 8-digit integer hash of the patient ID was subtracted from this value to encrypt the time of the examination without changing the chronological sequence of the examinations.

For example, a file name might look like *Figure 4*:

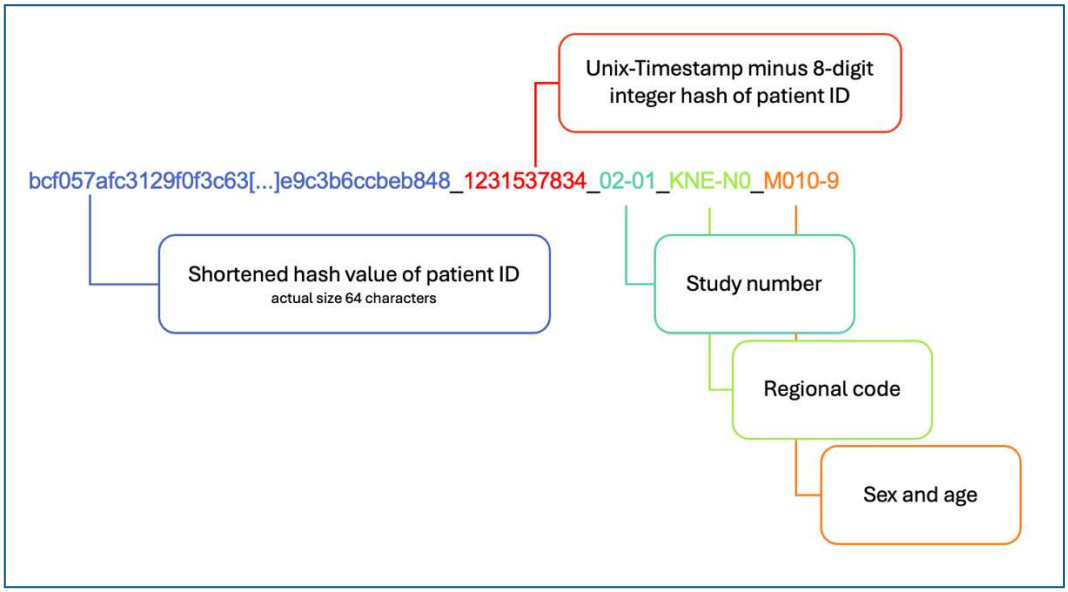

Figure 4 – Example of a file name

5.3 EfficientNet and training

EfficientNet is a family of CNNs that was developed by *Tan & Le* (113) in 2019 and is characterized by its resource-efficient architecture. The authors used so called ‘compound scaling’, a method in which all three dimensions of a network (depth, width, and resolution) are adjusted in a balanced manner at a fixed ratio, rather than being adjusted individually. The depth of a network refers to the number of layers and the width to the number of channels per layer. (83)

This approach was first implemented in the base model EfficientNet-B0 and served as the basis for deriving increasingly complex variants (B1 – B7). (113)

The neural networks were trained using a cross-validation methodology.

Specifically, 10-fold cross-validation was performed. In this procedure, the dataset is partitioned into 10 subsets (or 'folds'). The model is then trained 10 times; in each iteration, a different fold is held out for testing, while the remaining 9 folds are

used for training. The final performance metrics were calculated by averaging the results from all 10 folds. A key advantage of cross-validation is its ability to provide a robust performance evaluation even with a dataset that is relatively small for machine learning purposes, such as the one used in this study. This allows for a reliable assessment of a concept's potential viability.

5.4 Test parameters

Various test parameters are used to test the performance of a CNN. The most commonly used parameters are displayed in a *confusion matrix*, in which a comparison between predicted and actual classes can be made. A brief overview is shown in *Figure 5*.

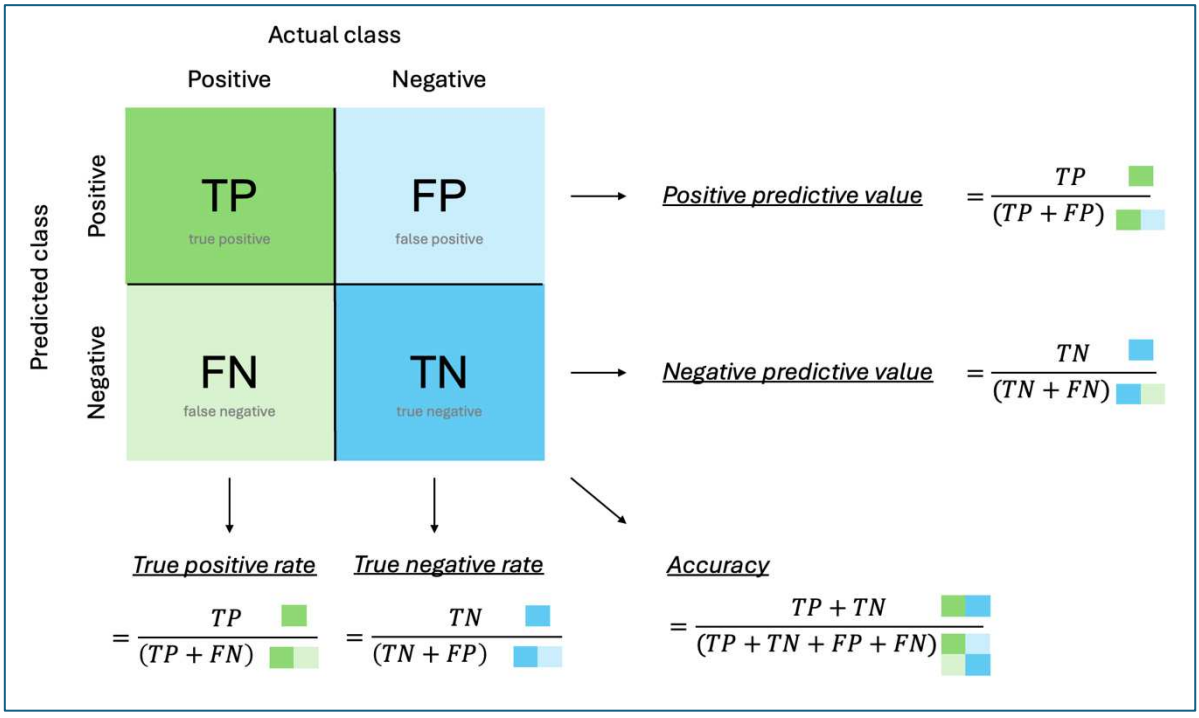

Figure 5 – Confusion matrix

When data is evaluated according to a binary classification system, a property can be either present or absent and a positive or negative value is assigned accordingly. In our case, the binaries are malignant (positive) and non-malignant (negative). If a case is correctly recognized as positive, corresponding to the actual class, it is considered a true positive (TP). A case that is classified as positive whereas the actual class is negative is considered a false positive (FP). Vise versa, negative results can be true negative (TN) or false negative (FN).

Different rates and values can be calculated with these four values. (114)

True positive rate:

The true positive rate (TPR) describes all values recognized as positive compared to all actual class positive values. It is a metric that can be used to estimate how many positive cases are recognized as positive. Other names for TPR are sensitivity and recall.

$$True\ positive\ rate\ (TPR) = Sensitivity = Recall = \frac{TP}{(TP + FN)}$$

True negative rate:

The true negative rate (TNR) is the counterpart to the TPR and describes the portion of actual negative cases that were true negatives. It is also called specificity.

$$True\ negative\ rate\ (TNR) = Specificity = \frac{TN}{(TN + FP)}$$

Positive and negative predictive value:

Both aforementioned rates have the disadvantage that in clinical practice the actual class of a case is often not known. Therefore, the positive predictive value (PPV) and negative predictive value (NPV) are used to make an assumption about the likelihood of a correct positive or negative result. The PPV can also be called precision

$$Positive\ predictive\ value\ (PPV) = Precision = \frac{TP}{(TP + FP)}$$

$$Negative\ predictive\ value\ (NPV) = \frac{TN}{(TN + FN)}$$

Accuracy:

Accuracy describes the fraction of correctly classified cases out of all cases.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

All of the previously mentioned metrics can be affected by class imbalance, meaning a class is considerably underrepresented, therefore skewing the results. For example, in distributions highly favoring negative cases, a few false positives can substantially reduce the precision as positive cases are rare overall. (115)

F1-score:

To better evaluate the performance of a model in a scenario with imbalanced class distribution, the F1-score is used. It represents the harmonic mean of precision (PPV) and recall (TPR), offering a balanced metric that takes both false positives and false negatives into account.

$$F1\text{-}score = \frac{2 * TP}{(2 * TP + FP + FN)} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

PR- and ROC-curves

Precision-Recall (PR) curves plot precision and recall against each other.

Receiver Operating Characteristic (ROC) shows the discriminatory power of a model by plotting true positive rate and false positive rate against each other.

5.5 Statistical analysis

Descriptive statistics were used to summarize the respective metrics. As no comparison groups were available, inferential statistical tests were omitted. All values were calculated using IBM SPSS Statistics v.27 (IBM, Armonk, NY, USA).

6. Results

Tumor localizations were categorized into five major anatomical groups. A detailed breakdown of the distribution within these groups, stratified by tumor dignity, is presented in *Table 5*. Notably, the group of lesions in the axial skeleton (skull, spine, and ribs) exhibited a markedly higher proportion of malignant tumors compared to all other localization groups.

Localization	Benign	Malignant	Overall
Lower extremity	376 [92.67%]	30 [7.39%]	406
Upper extremity	234 [94.74%]	13 [5.26%]	247
Pelvis	48 [82.76%]	10 [17.24%]	58
Axial skeleton	13 [27.66%]	34 [72.34%]	47
Shoulder girdle	36 [85.71%]	6 [14.29%]	42
Overall	**707 [88.37%]**	**93 [11.62%]**	**800**

Table 5 – Tumor localizations

The performance of the EfficientNet variants (B0 – B7) is shown in *Table 6*, *Table 7*, and *Table 8* based on accuracy, precision, recall, and F1-score. *Table 6* and *Table 7* show the performance for benign and malignant lesions, respectively. *Table 8* shows the overall performance and additionally the values of PR-AUC (area under the curve) and ROC-AUC. The curves themselves are presented in *Figure 6* and *Figure 7.*

Performance benign			
Modell	**Precision**	**Recall**	**F1-score**
B0	0.911	0.986	0.947
B1	0.910	0.999	0.952
B2	0.912	**1.000**	**0.954**
B3	*0.909*	0.999	0.951
B4	0.912	0.996	0.952
B5	0.912	0.993	0.951
B6	0.913	0.984	0.948

B7	**0.914**	*0.979*	*0.945*

Table 6 – Performance of EfficientNet B0 to B7 on benign lesions, highest value bold, lowest value cursive

All models achieved high metrics in the classification of benign lesions. Recall was 1.000 for B2 and nearly as high for all other variants (between 0.979 and 0.999). Precision was slightly lower, ranging from 0.909 to 0.914, with B7 showing the highest precision. As the harmonic mean between precision and recall, the F1-score was highest in B2 at 0.954. All other variants showed values in a similar range (0.945 – 0.952).

Performance malignant

Modell	**Precision**	**Recall**	**F1-score**
B0	0.714	0.269	0.391
B1	0.958	0.247	0.393
B2	**1.000**	0.269	**0.424**
B3	0.957	*0.237*	*0.379*
B4	0.893	0.269	0.413
B5	0.833	0.269	0.407
B6	0.711	0.290	0.412
B7	*0.651*	**0.301**	0.412

Table 7 – Performance of EfficientNet B0 – B7 on malignant lesions, highest value bold, lowest value cursive

The results for malignant lesions were notably worse overall. Recall values were consistently low, ranging from 0.237 to 0.301. Although precision showed better values, they were widely scattered: 0.651 for B7 and 1.000 for B2. The performance is reflected in the low F1-scores, ranging from 0.379 to 0.424, with B2 showing the highest score once again.

Overall performance

Modell	Accuracy	Precision	Recall	F1-score	PR-AUC	ROC-AUC
B0	0.903	0.813	0.627	0.*669*	0.466	0.677
B1	0.911	0.934	0.623	0.673	0.620	0.662
B2	**0.915**	**0.956**	0.634	**0.689**	0.617	0.657
B3	0.910	0.933	*0.618*	0.665	**0.624**	*0.642*
B4	0.911	0.902	0.632	0.683	0.591	0.669
B5	0.909	0.873	0.631	0.679	0.616	0.686
B6	0.904	0.812	0.637	0.680	0.483	0.664
B7	*0.900*	*0.783*	**0.640**	0.679	*0.450*	**0.694**

Table 8 – Overall performance of EfficientNet B0 – B7, highest value bold, lowest value cursive

In terms of overall performance, EfficientNet B2 demonstrated the greatest accuracy with 0.915. This variant also showed the highest precision with 0.956, as well as the highest F1-score with 0.689. B7 achieved the highest recall at 0.640, although the corresponding F1-score is lower than that of B2.

B3 achieves the highest PR-AUC with 0.624, which is similar to the values of B1 (0.620), B2 (0.617) and B5 (0.616). B7 shows the weakest performance with 0.450, with B0 and B6 showing similar values at 0.466 and 0.483, respectively

As shown in *Figure 6*, the PR-curves show some differences in the performance of all models but are generally flat.

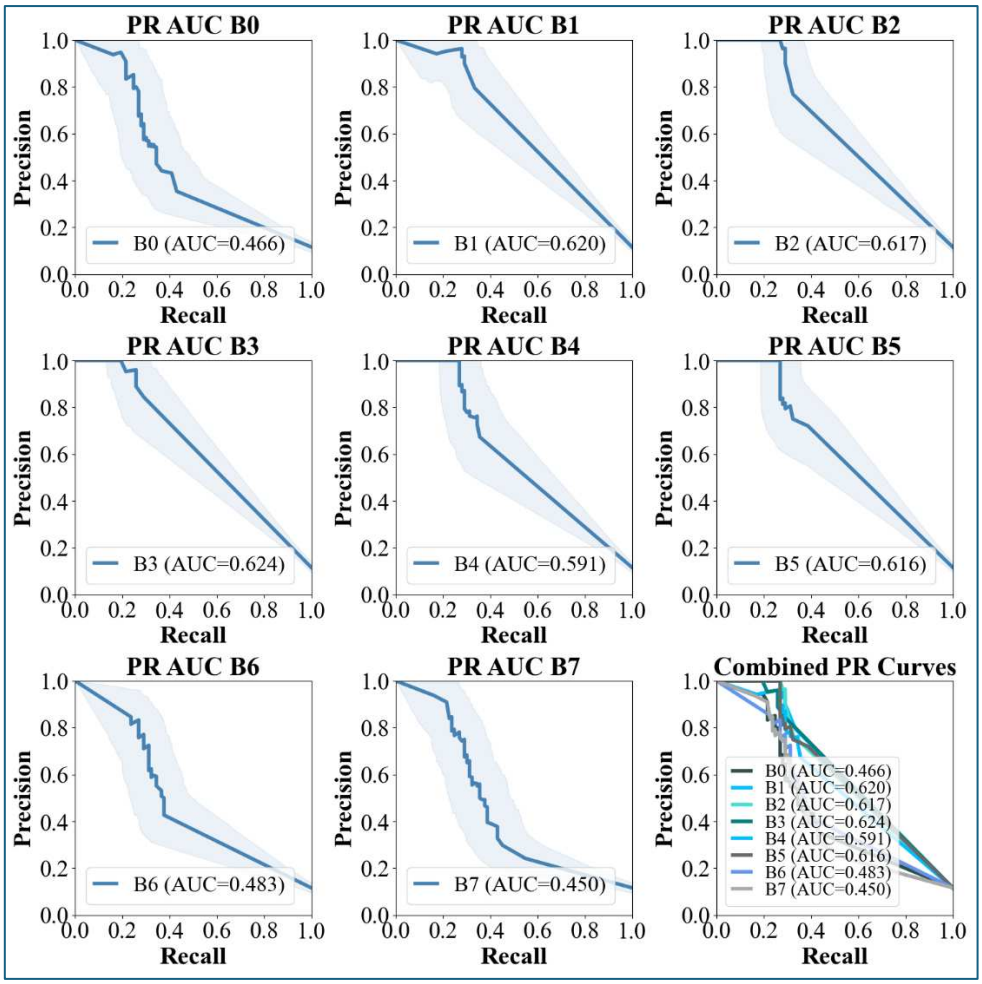

Figure 6 – PR-curves with calculated AUC for models B0 – B7 and all models combined

Figure 7 depicts the ROC-curves, which plots the recall (TPR) against the FPR, with their calculated AUCs.

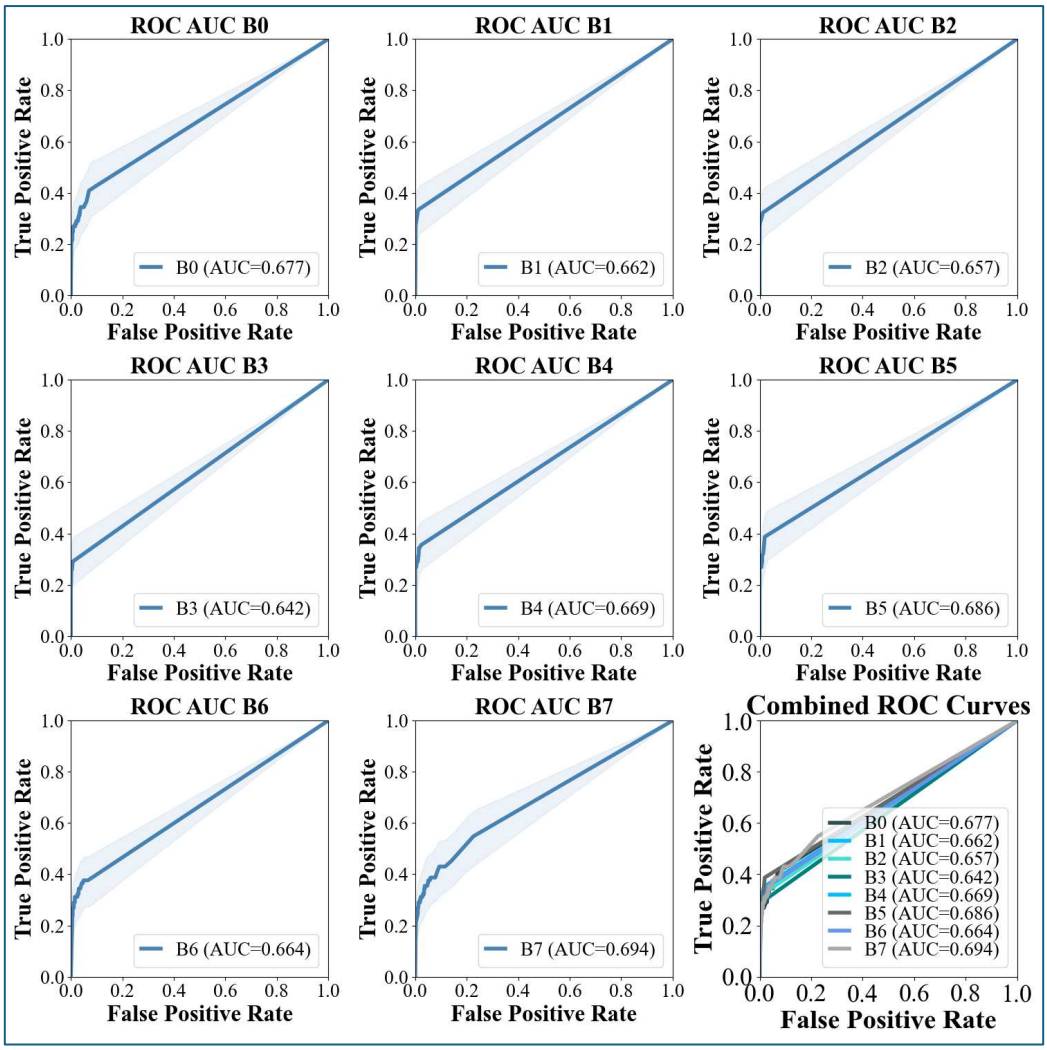

Figure 7 – ROC-AUC curves for Model B0 – B7 and all models combined

The ROC-AUC values are closer together, ranging from 0.642 to 0.694. B7 shows the highest value, and B3 the lowest.

Figure 8 shows the summed confusion matrix for all models. *Figure 9* and *Figure 10* show the confusion matrices for all individual EfficientNet models B0 to B7.

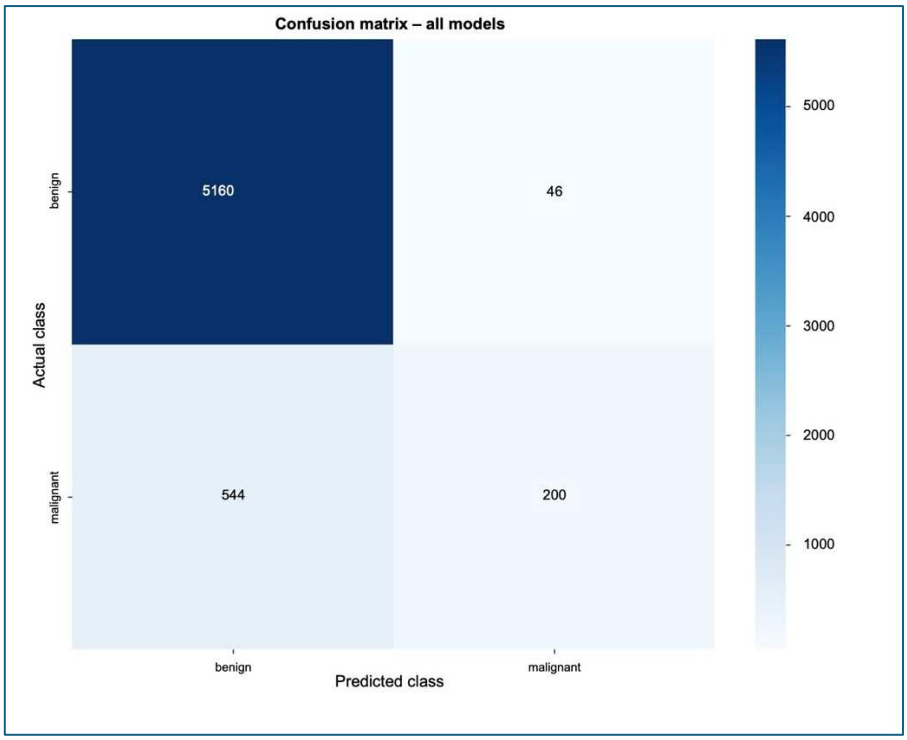

Figure 8 – Confusion matrix all models

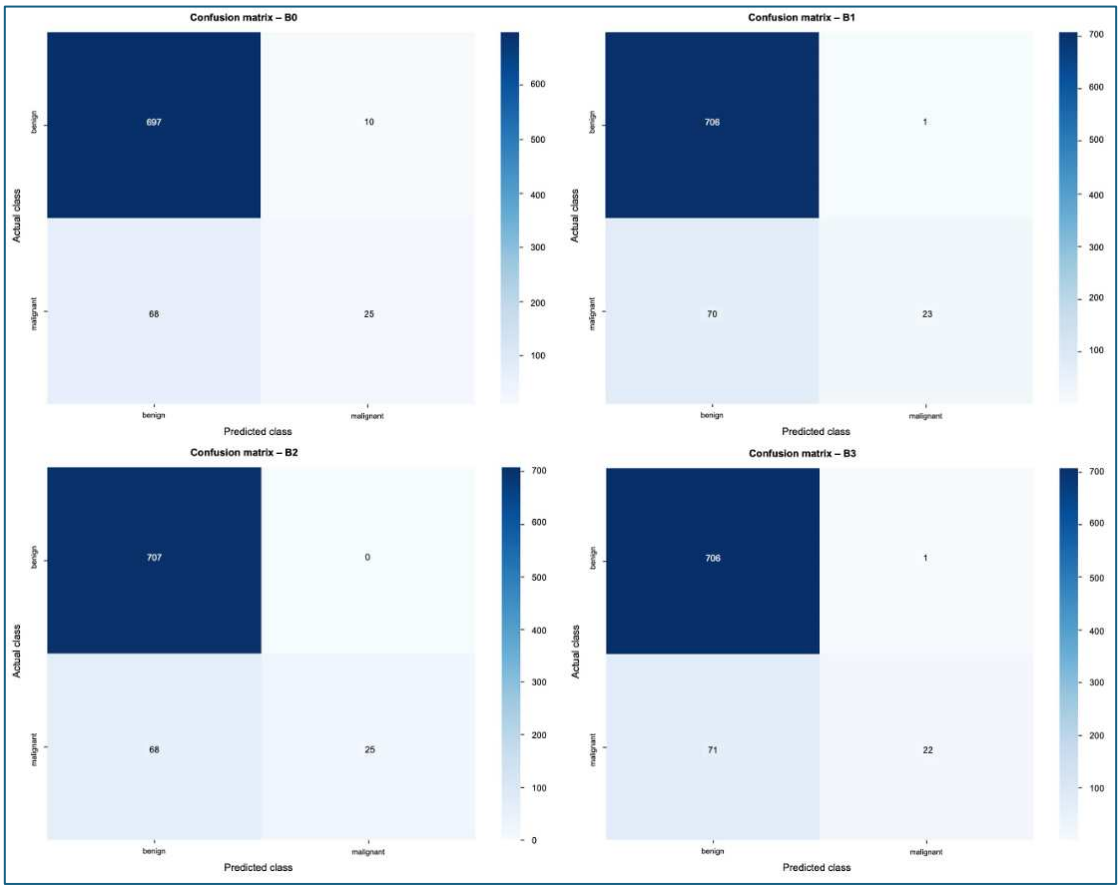

Figure 9 – Confusion matrix model B0, B1, B2, B3

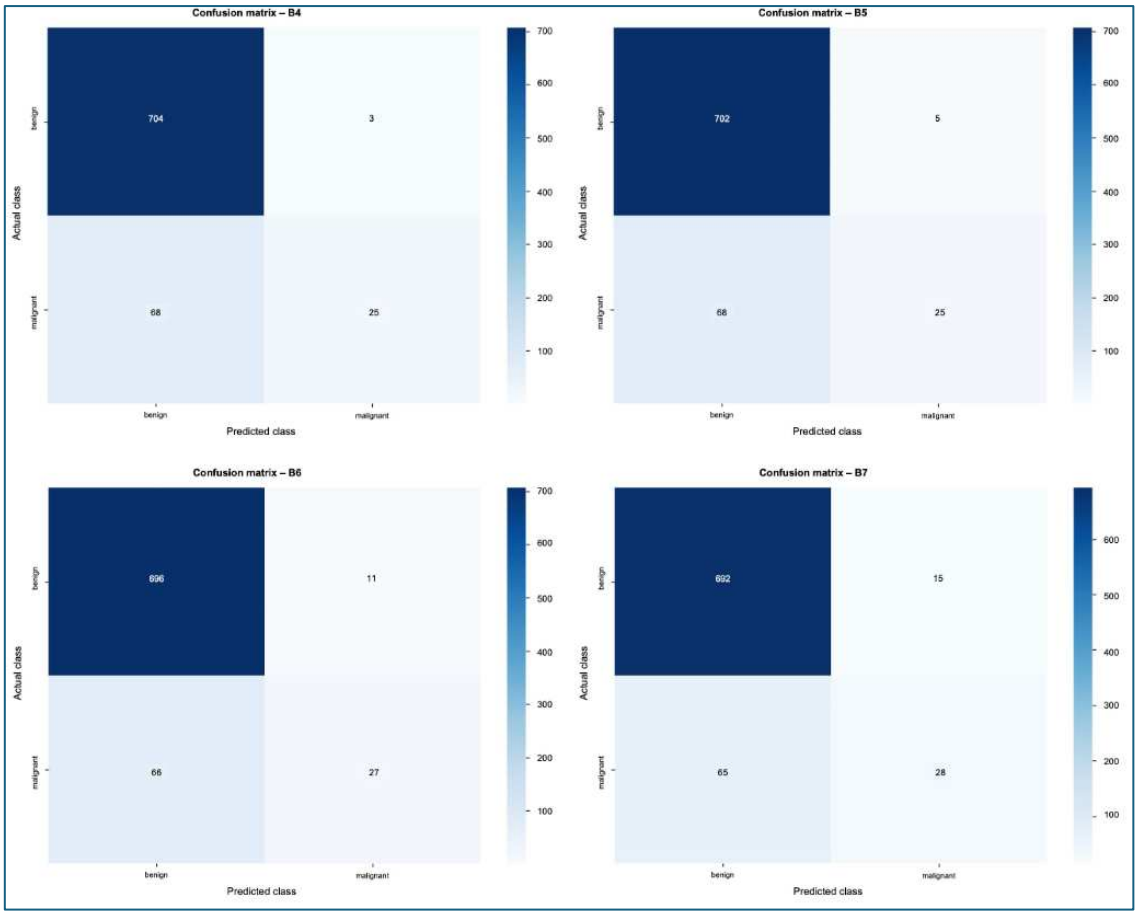

Figure 10 – Confusion matrix model B4, B5, B6, B7

To visualize the model's decision process, heatmaps were generated using Grad-Cam++. *Figure 11* shows two example heatmaps of a benign and a malignant case from the same patient. The patient of the benign case was a 4.6-year-old male, the patient of the malignant case a 14.6-year-old male.

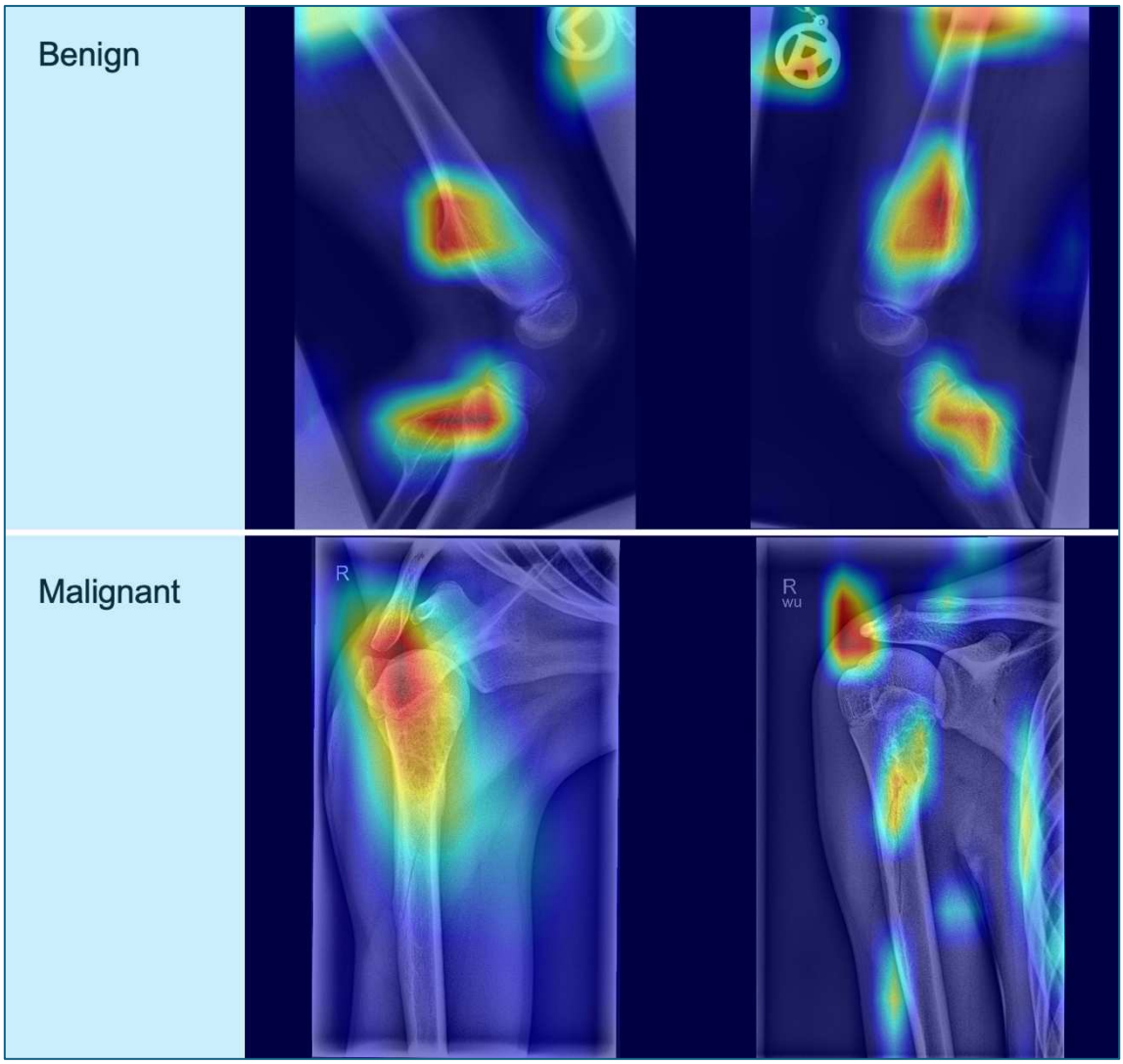

Figure 11 – Visualization using Grad-CAM++. Upper row: benign, male, 4.6 years of age. Lower row: malignant, male, 14.6 years of age.

7. Discussion

For this thesis, we trained and tested eight different variants of the CNN architecture EfficientNet for the classification of benign and malignant primary bone tumors in children. The results show that EfficientNet models are well-suited for the classification of benign tumors. This is reflected in the high overall recall for the benign group, with EfficientNet-B2 reaching a recall as high as 1.00, meaning all benign tumors were correctly identified as such. All other models showed similar result with lowest performance from B7 at 0.979. The precision of the benign classification scored lower, ranging from 0.909 (B3) to 0.914 (B7), meaning

90 – 91% of tumors classified as benign were indeed benign. However, this also means that about 10% of tumors labeled as benign were actually malignant.

The classification of malignant tumors did not yield satisfactory result. The recall showed that, on average, only 27% of malignant tumors were identified as such. Precision values ranged between a minimum of 0.651 (B7) to a maximum of 1.000 (B2). This indicated that while B2 was able to correctly assign a malignant label to actually malignant lesions, it did so by missing over 70% of malignant cases.

The varying results for benign and malignant tumors are reflected in the overall performance results. The combination of high precision (on average 0.876) and low recall (on average 0.630) suggests that the models only reported a malignant lesion where they were highly confident, as shown in the low false positive rates for malignancy. Consequently, many true malignant cases were overlooked, resulting in a high false negative rate for malignancy. The most likely contributing factor to this problem is the class imbalance in the dataset. With a ratio of approximately 88% benign to 12% malignant lesions, benign findings far outweighed malignant findings in the training material. This imbalance makes reliably detecting malignant tumors considerably more difficult.

Due to the imbalance in the material, the high accuracy of on average 91% in the overall performance should not be overinterpreted, since accuracy is easily distorted by class imbalance. A more reliable and accurate marker is the F1 score as a harmonic mean between recall and precision, which averages on 0.677 in the overall performance.

Other reliable markers for imbalanced datasets are PR curves and their corresponding AUCs. The highest PR-AUC was achieved by B3 (0.624), the lowest by B7 (0.450). For the ROC curves, all models are very close in performance, with a range from 0.624 (B3) to 0.694 (B7). The diagonal running from the bottom left corner to the top right of the ROC plot, is called the ‘line of no-discrimination’, noted with a value of 0.5 and considered to be equivalated to a random guess. Anything value above 0.5 indicated that the model works better than random. While most models’ curves showed promising initial growth, their performance dropped off quickly and closely followed the diagonal.

When comparing all model variants, EfficientNet-B2 shows the highest F1 score (0.689), the highest precision (0.956), and an above average recall (0.634) resulting in a relatively high PR-AUC (0.617). The highest recall was achieved by B7 (0.640), as well as the highest ROC-AUC (0.694), while the F1 score (0.679), precision (0.783) and PR-AUC (0.450) were among the lowest.

In an oncological context, a false negative (missed diagnosis) must be avoided at all costs, since it can be detrimental to the patient's prognosis. This is especially important in the pediatric context. A false positive can also cause significant harm, anxiety, and unnecessary stress for the patient and their caregivers. The models trained in this study are currently not ready to provide a reliable diagnosis.

Our study was performed at a single-center institution with data that was collected retrospectively, thus creating a selection bias and putting the generalizability of the model's performance into question. No external test set was used, further restricting the validity of the results.

The main limiting factor is the pronounced class imbalance between benign and malignant cases. In addition, another factor made the identification of malignant lesions more difficult. More than a third of malignant cases (34 out of 93) cases occurred in the axial skeleton, which is a much more complex anatomical area than, for example, long bones, making the assessment of these lesions more challenging, which might have contributed to the reduced model performance in malignant lesions. A stratified subgroup analysis according to anatomical location would be a possible next step towards more accurate characterization of the models' performances.

While the distribution of cases is a significant limitation to this study, it more accurately represents clinical reality, in which benign tumors heavily outweigh malignant tumors. We included both histopathologically confirmed diagnoses and lesion with a clear radiological appearance, such as NOFs. The main limitation of this approach is of course the reinforcement of class imbalance. It also means that, in some cases, the 'ground truth' is solely based on radiological assessment.

From a technical standpoint, however, the study offers a practical approach. Unlike many previous studies in which image section were manually cropped or

annotated, we applied the model to unmodified DICOM images exported as PNG files. This end-to-end approach is the first attempt at integrating an AI system into a clinically realistic environment without manual preprocessing.

A common clinical question is the differentiation between tumors and diseases that might appear similar to tumors, like infections. As this was not the aim of this study, differential diagnoses were not included but have been addressed by other authors. (99,111)

To our knowledge, we created the largest exclusively pediatric dataset of primary bone tumors to date. The age distribution spans the entire age range from newborns to young adults. While this demographic spread strengthens the dataset's representativeness, it may also pose a limitation, as the radiographic appearance of bones changes drastically during development and growths.

Despite the relatively large number of cases, the cohort sizes of the different entities remain moderate in the context of training deep neural networks. We believe that this dataset serves as a very useful starting ground for further research, as future studies could use generative adversarial networks (GANs) to synthetically increase the database size. Additionally, pre-trained models could be of use.

Notably, a comprehensive, exclusively pediatric dataset was created, providing a valuable basis for future research projects. Stratification by anatomical region (e.g. axial vs. appendicular), age group or tumor entities, is easily possible. In addition, the models' the performances for malignant lesions could be significantly improved in the future, if sufficient training data is generated or collected.

VIII. References

1. Mehta K, McBee MP, Mihal DC, England EB. Radiographic Analysis of Bone Tumors: A Systematic Approach. Semin Roentgenol. 2017 Oct 1;52(4):194–208.

2. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin [Internet]. 2025 Jan 16 [cited 2025 May 26]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/39817679

3. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001–10: a population-based registry study. Lancet Oncol. 2017 Jun 1;18(6):719–31.

4. Kaatsch P, Strothotte J, Becker C, Bielack S, Dirksen U, Blettner M. Pediatric bone tumors in Germany from 1987 to 2011: incidence rates, time trends and survival. Acta Oncol (Madr) [Internet]. 2016 Oct 2 [cited 2025 May 26];55(9–10):1145–51. Available from: https://medicaljournalssweden.se/actaoncologica/article/view/25550

5. Cole S, Gianferante DM, Zhu B, Mirabello L. Osteosarcoma: A Surveillance, Epidemiology, and End Results program-based analysis from 1975 to 2017. Cancer [Internet]. 2022 Jun 1 [cited 2025 Jun 22];128(11):2107. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11647566/

6. Spector LG, Hubbard AK, Diessner BJ, Machiela MJ, Webber BR, Schiffman JD. Comparative international incidence of Ewing sarcoma 1988 to 2012. Int J Cancer. 2021 Sep 1;149(5):1054–66.

7. Breden S, Stephan M, Hinterwimmer F, Consalvo S, Lenze U, von Eisenhart-Rothe R, et al. Pediatric Bone Tumors: Location and Age Distribution of 420 Cases. Diagnostics [Internet]. 2024 Nov 1 [cited 2025 May 26];14(22):2513. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11593068/

8. Yildiz C, Erler K, Atesalp AS, Basbozkurt M. Benign bone tumors in children. Curr Opin Pediatr. 2003 Feb;15(1):58–67.

9. Ritter J, Bielack SS. Osteosarcoma. Annals of Oncology. 2010 Oct;21(SUPPL. 7).

10. Nichols RE, Dixon LB. Radiographic Analysis of Solitary Bone Lesions. Radiol Clin North Am [Internet]. 2011 Nov 1 [cited 2025 Jun 5];49(6):1095–114. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0033838911001333

11. Sadykova LR, Ntekim AI, Muyangwa-Semenova M, Rutland CS, Jeyapalan JN, Blatt N, et al. Epidemiology and Risk Factors of Osteosarcoma. Cancer Invest. 2020 May 27;38(5):259–69.

12. Hardes J, Gosheger G, Budny T. Knochensarkome. Z Orthop Unfall [Internet]. 2018 Feb 1 [cited 2025 May 31];156(1):105–24. Available from: https://pubmed.ncbi.nlm.nih.gov/29471560/

13. Arora RS, Kontopantelis E, Alston RD, Eden TO, Geraci M, Birch JM. Relationship between height at diagnosis and bone tumours in young people: A meta-analysis. Cancer Causes and Control. 2011 May;22(5):681–8.

14. Balamuth NJ, Womer RB. Ewing's sarcoma. Lancet Oncol [Internet]. 2010 Feb 1 [cited 2025 May 31];11(2):184–92. Available from: https://www.thelancet.com/action/showFullText?pii=S1470204509702864

15. Schulte M, Hartmann W. Bone tumors in children and adolescents. Pathologie. 2023 Nov 1;44(6):348–56.

16. De Salvo S, Pavone V, Coco S, Dell'agli E, Blatti C, Testa G. Benign Bone Tumors: An Overview of What We Know Today. Journal of Clinical Medicine 2022, Vol 11, Page 699 [Internet]. 2022 Jan 28 [cited 2025 Jun 10];11(3):699. Available from: https://www.mdpi.com/2077-0383/11/3/699/htm

17. Girish H, Mittal P, Aravind M. Non-ossifying fibroma of distal tibia: case report. Int J Res Orthop [Internet]. 2024 Dec 29 [cited 2025 Jun

4];10(1):216–8. Available from: https://www.ijoro.org/index.php/ijoro/article/view/3010

18. Motamedi K, Seeger LL. Benign bone tumors. Radiol Clin North Am [Internet]. 2011 Nov 1 [cited 2025 Jun 2];49(6):1115–34. Available from: https://escholarship.org/uc/item/7h86k14t

19. Pullan JE, Lotfollahzadeh S. Primary Bone Cancer. Nurs Stand [Internet]. 2024 Mar 20 [cited 2025 May 26];28(12):22. Available from: https://www.ncbi.nlm.nih.gov/books/NBK560830/

20. Canavese F, Samba A, Rousset M. Pathological fractures in children: Diagnosis and treatment options. Orthopaedics & Traumatology: Surgery & Research [Internet]. 2016 Feb 1 [cited 2025 Jun 21];102(1):S149–59. Available from: https://www.sciencedirect.com/science/article/pii/S1877056815003461?via%3Dihub

21. Siegal GP, Bloem JL, Cates JMM HM. Soft Tissue and Bone Tumours. Vol. 3, WHO Iarc. WORLD HEALTH ORGANIZATION; 2020. 472–474 p.

22. Choi JH, Ro JY. The 2020 WHO Classification of Tumors of Bone: An Updated Review. Adv Anat Pathol [Internet]. 2021 May 1 [cited 2025 Jun 7];28(3):119–38. Available from: https://pubmed.ncbi.nlm.nih.gov/33480599/

23. Hwang S, Hameed M, Kransdorf M. The 2020 World Health Organization classification of bone tumors: what radiologists should know. Skeletal Radiol [Internet]. 2023 Mar 1 [cited 2025 Jun 7];52(3):329–48. Available from: https://link.springer.com/article/10.1007/s00256-022-04093-7

24. Baumhoer D, Hench J, Amary F. Recent advances in molecular profiling of bone and soft tissue tumors. Skeletal Radiol [Internet]. 2024 Sep 1 [cited 2025 Jun 22];53(9):1925–36. Available from: https://link.springer.com/article/10.1007/s00256-024-04584-9

25. WHO Classification of Tumours Editorial Board. Paediatric Tumours. International Agency for Research on Cancer, World Health Organization; 2022. 1191 p.

26. Pfister SM, Reyes-Múgica M, Chan JKC, Hasle H, Lazar AJ, Rossi S, et al. A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. Cancer Discov [Internet]. 2021 Feb 1 [cited 2025 Jun 7];12(2):331. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9401511/

27. Enneking WF, Spanier SS, Goodman M. A system for the surgical staging of musculoskeletal sarcoma. Clin Orthop Relat Res. 1980;153:106–20.

28. ENNEKING WF. A System of Staging Musculoskeletal Neoplasms. Clin Orthop Relat Res [Internet]. 1986;204. Available from: https://journals.lww.com/clinorthop/fulltext/1986/03000/a_system_of_staging_musculoskeletal_neoplasms.3.aspx

29. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging . CA Cancer J Clin [Internet]. 2017 Mar 1 [cited 2025 Jun 22];67(2):93–9. Available from: /doi/pdf/10.3322/caac.21388

30. Widhe B, Widhe T. Initial symptoms and clinical features in osteosarcoma and Ewing sarcoma. Journal of Bone and Joint Surgery. 2000;82(5):667–74.

31. Heare T, Hensley MA, Dell’Orfano S. Bone tumors: Osteosarcoma and Ewing’s sarcoma. Curr Opin Pediatr [Internet]. 2009 Jun [cited 2025 May 31];21(3):365–72. Available from: https://journals.lww.com/co-pediatrics/fulltext/2009/06000/bone_tumors__osteosarcoma_and_ewing_s_sarcoma.16.aspx

32. Longhi A, Errani C, De Paolis M, Mercuri M, Bacci G. Primary bone osteosarcoma in the pediatric age: State of the art. Cancer Treat Rev. 2006 Oct;32(6):423–36.

33. Choi EYK, Gardner JM, Lucas DR, McHugh JB, Patel RM. Ewing sarcoma. Semin Diagn Pathol [Internet]. 2014 [cited 2025 Jun 22];31(1):39–47. Available from: https://pubmed.ncbi.nlm.nih.gov/24680181/

34. Herget GW, Mauer D, Krauß T, El Tayeh A, Uhl M, Südkamp NP, et al. Non-ossifying fibroma: Natural history with an emphasis on a stage-related growth, fracture risk and the need for follow-up. BMC Musculoskelet Disord [Internet]. 2016 Apr 5 [cited 2025 Jun 4];17(1):1–7. Available from: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-016-1004-0

35. Tepelenis K, Papathanakos G, Kitsouli A, Troupis T, Barbouti A, Vlachos K, et al. Osteochondromas: An Updated Review of Epidemiology, Pathogenesis, Clinical Presentation, Radiological Features and Treatment Options. In Vivo (Brooklyn) [Internet]. 2021 Mar 1 [cited 2025 Jun 10];35(2):681. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8045119/

36. Restrepo R, Zahrah D, Pelaez L, Temple HT, Murakami JW. Update on aneurysmal bone cyst: pathophysiology, histology, imaging and treatment. Pediatr Radiol [Internet]. 2022 Aug 1 [cited 2025 Jun 21];52(9):1601–14. Available from: https://pubmed.ncbi.nlm.nih.gov/35941207/

37. Parmeggiani A, Miceli M, Errani C, Facchini G, Palmerini E, Schwab J, et al. State of the Art and New Concepts in Giant Cell Tumor of Bone: Imaging Features and Tumor Characteristics. Cancers 2021, Vol 13, Page 6298 [Internet]. 2021 Dec 15 [cited 2025 Jun 8];13(24):6298. Available from: https://www.mdpi.com/2072-6694/13/24/6298/htm

38. Salotti JA, Nanduri V, Pearce MS, Parker L, Lynn R, Windebank KP. Incidence and clinical features of Langerhans cell histiocytosis in the UK and Ireland. Arch Dis Child [Internet]. 2009 May 1 [cited 2025 Jul 30];94(5):376–80. Available from: https://adc.bmj.com/content/94/5/376

39. Copley L, Dormans JP. Benign pediatric bone tumors: Evaluation and treatment. Pediatr Clin North Am. 1996;43(4):949–66.

40. Murphey MD, Choi JJ, Kransdorf MJ, Flemming DJ, Gannon FH. From the archives of the AFIP. Imaging of osteochondroma: Variants and complications with radiologic-pathologic correlation. Radiographics [Internet]. 2000 Sep 1 [cited 2025 Jun 10];20(5):1407–34. Available from: /doi/pdf/10.1148/radiographics.20.5.g00se171407

41. Garcia SA, Ng VY, Iwamoto M, Enomoto-Iwamoto M. Osteochondroma Pathogenesis: Mouse Models and Mechanistic Insights into Interactions with Retinoid Signaling. Am J Pathol [Internet]. 2021 Dec 1 [cited 2025 Jun 10];191(12):2042. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8647428/

42. Verdegaal SHM, Bovée JVMG, Pansuriya TC, Grimer RJ, Ozger H, Jutte PC, et al. Incidence, Predictive Factors, and Prognosis of Chondrosarcoma in Patients with Ollier Disease and Maffucci Syndrome: An International Multicenter Study of 161 Patients. Oncologist [Internet]. 2011 Dec 1 [cited 2025 Jun 20];16(12):1771–9. Available from: https://dx.doi.org/10.1634/theoncologist.2011-0200

43. Oliveira AM, Perez-Atayde AR, Inwards CY, Medeiros F, Derr V, Hsi BL, et al. USP6 and CDH11 oncogenes identify the neoplastic cell in primary aneurysmal bone cysts and are absent in so-called secondary aneurysmal bone cysts. American Journal of Pathology [Internet]. 2004 Nov 1 [cited 2025 Jun 22];165(5):1773–80. Available from: https://ajp.amjpathol.org/action/showFullText?pii=S0002944010634323

44. Dabska M, Buraczewski J. Aneurysmal bone cyst. Pathology, clinical course and radiologic appearances. Cancer [Internet]. 1969 Feb 1 [cited 2025 Jun 22];23(2):371–89. Available from: /doi/pdf/10.1002/1097-0142%28196902%2923%3A2%3C371%3A%3AAID-CNCR2820230213%3E3.0.CO%3B2-2

45. Mascard E, Gomez-Brouchet A, Lambot K. Bone cysts: Unicameral and aneurysmal bone cyst. Orthopaedics and Traumatology: Surgery and Research. 2015 Feb 1;101(1):S119–27.

46. Zishan US, Pressney I, Khoo M, Saifuddin A. The differentiation between aneurysmal bone cyst and telangiectatic osteosarcoma: a clinical, radiographic and MRI study. Skeletal Radiol. 2020 Sep 1;49(9):1375–86.

47. Phemister DB. CHRONIC FIBROUS OSTEOMYELITIS. Ann Surg [Internet]. 1929 Oct [cited 2025 Jun 5];90(4):756. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1399079/

48. Sontag LW, Pyle SI. The appearance and nature of cyst-like areas in the distal femoral metaphyses of children. Am J Roentgenol. 1941;46(2):185–8.

49. Jaffe HL, Lichtenstein L. Non-osteogenic fibroma of bone. Am J Pathol [Internet]. 1942 Mar [cited 2025 Jun 5];18(2):205. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2032933/

50. Bowers LM, Cohen DM, Bhattacharyya I, Pettigrew JC, Stavropoulos MF. The Non-ossifying Fibroma: A Case Report and Review of the Literature. Head Neck Pathol [Internet]. 2012 Jun [cited 2025 Jun 4];7(2):203. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3642261/

51. Arata MA, Peterson HA, Dahlin DC. Pathological fractures through non-ossifying fibromas. Review of the Mayo Clinic experience. JBJS [Internet]. 1981;63(6). Available from: https://journals.lww.com/jbjsjournal/fulltext/1981/63060/pathological_fractures_through_non_ossifying.16.aspx

52. Temple HT, Scully SP, Aboulafia AJ. Benign bone tumors. Instr Course Lect [Internet]. 2002 [cited 2025 Jun 5];51:429–39. Available from: https://pure.johnshopkins.edu/en/publications/benign-bone-tumors

53. Dumitriu DI, Menten R, Clapuyt P. Pitfalls in the diagnosis of common benign bone tumours in children. Insights Imaging [Internet]. 2014 Dec 12 [cited 2025 Jun 8];5(6):645–55. Available from: https://insightsimaging.springeropen.com/articles/10.1007/s13244-014-0356-y

54. Rechl H, Kirchhoff C, Wörtler K, Lenze U, Töpfer A, Von Eisenhart-Rothe R. Diagnostik von malignen Knochen- und Weichteiltumoren. Orthopade. 2011 Oct;40(10):931–44.

55. Luedtke LM, Flynn JM, Ganley TJ, Hosalkar HS, Pill SG, Dormans JP. The orthopedists' perspective: Bone tumors, scoliosis, and trauma. Radiol Clin North Am [Internet]. 2001 [cited 2025 Jun 5];39(4):803–21. Available from: https://pubmed.ncbi.nlm.nih.gov/11549172/

56. Voss SD, Reaman GH, Kaste SC, Slovis TL. The ALARA concept in pediatric oncology. Pediatr Radiol [Internet]. 2009 Nov 30 [cited 2025 Jun 20];39(11):1142–6. Available from: https://link.springer.com/article/10.1007/s00247-009-1404-5

57. Hall EJ. Radiation biology for pediatric radiologists. Pediatr Radiol [Internet]. 2009 Feb 16 [cited 2025 Jun 20];39(SUPPL. 1):57–64. Available from: https://link.springer.com/article/10.1007/s00247-008-1027-2

58. Cederberg KB, Iyer RS, Chaturvedi A, McCarville MB, McDaniel JD, Sandberg JK, et al. Imaging of pediatric bone tumors: A COG Diagnostic Imaging Committee/SPR Oncology Committee White Paper. Pediatr Blood Cancer. 2023 Jun 1;70(S4).

59. Kaste SC. Imaging pediatric bone sarcomas. Radiol Clin North Am. 2011 Jul;49(4):749–65.

60. Lodwick GS, Wilson AJ, Farrell C, Virtama P, Dittrich F. Determining growth rates of focal lesions of bone from radiographs. Radiology [Internet]. 1980 [cited 2025 Jun 19];134(3):577–83. Available from: https://pubmed.ncbi.nlm.nih.gov/6928321/

61. Crombé A, Simonetti M, Longhi A, Hauger O, Fadli D, Spinnato P. Imaging of Osteosarcoma: Presenting Findings, Metastatic Patterns, and Features Related to Prognosis. J Clin Med [Internet]. 2024 Oct 1 [cited 2025 Jun 18];13(19). Available from: https://pubmed.ncbi.nlm.nih.gov/39407770/

62. Weiss A, Khoury JD, Hoffer FA, Wu J, Billups CA, Heck RK, et al. Telangiectatic osteosarcoma: The St. Jude Children's Research Hospital's experience. Cancer [Internet]. 2007 Apr 15 [cited 2025 Jul 30];109(8):1627–37. Available from: /doi/pdf/10.1002/cncr.22574

63. McCarville MB, Chen JY, Coleman JL, Li Y, Li X, Adderson EE, et al. Journal club: Distinguishing osteomyelitis from ewing sarcoma on radiography and mri. American Journal of Roentgenology [Internet]. 2015 Sep 1 [cited 2025 Jul 30];205(3):640–51. Available from: /doi/pdf/10.2214/AJR.15.14341?download=true

64. Goyal N, Kalra M, Soni A, Baweja P, Ghonghe NP. Multi-modality imaging approach to bone tumors - State-of-the art. J Clin Orthop Trauma [Internet]. 2019 Jul 1 [cited 2025 Jul 30];10(4):687. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6611851/

65. Colliot O. A Non-technical Introduction to Machine Learning. Neuromethods. 2023;197:3–23.

66. Shortliffe EH. Books: Computer-Based Medical Consultations: MYCIN. J Clin Eng [Internet]. 1976;1(1). Available from: https://journals.lww.com/jcejournal/fulltext/1976/10000/books__computer_based_medical_consultations__mycin.11.aspx

67. Shen D, Wu G, Suk H Il. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng [Internet]. 2017 Jun 21 [cited 2025 Jul 31];19:221. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5479722/

68. Sogancioglu E, Çallı E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Med Image Anal [Internet]. 2021 Aug 1 [cited 2025 Jul 31];72. Available from: http://arxiv.org/abs/2103.08700

69. Rezayi S, R Niakan Kalhori S, Saeedi S. Effectiveness of Artificial Intelligence for Personalized Medicine in Neoplasms: A Systematic Review. Biomed Res Int [Internet]. 2022 [cited 2025 Jul 31];2022:7842566. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9010213/

70. Zhang Z, Wei X. Artificial intelligence-assisted selection and efficacy prediction of antineoplastic strategies for precision cancer therapy. Semin Cancer Biol [Internet]. 2023 May 1 [cited 2025 Jul 31];90:57–72. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1044579X23000287?via%3Dihub

71. Huhulea EN, Huang L, Eng S, Sumawi B, Huang A, Aifuwa E, et al. Artificial Intelligence Advancements in Oncology: A Review of Current Trends and Future Directions. Biomedicines 2025, Vol 13, Page 951 [Internet]. 2025 Apr 13 [cited 2025 Jul 31];13(4):951. Available from: https://www.mdpi.com/2227-9059/13/4/951/htm

72. Wang J, Zeng Z, Li Z, Liu G, Zhang S, Luo C, et al. The clinical application of artificial intelligence in cancer precision treatment. Journal of Translational Medicine 2025 23:1 [Internet]. 2025 Jan 27 [cited 2025 Jul 31];23(1):1–16. Available from: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-025-06139-5

73. Liu J. AI-Based Epidemic Spread Prediction and Public Health Response Optimization: A Systematic Study from Data Analysis to Policy Implementation. Applied and Computational Engineering [Internet]. 2024 Dec 19 [cited 2025 Jul 31];118(1):1–7. Available from: https://www.ewadirect.com/proceedings/ace/article/view/18470

74. Villanueva-Miranda I, Xiao G, Xie Y. Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review. Front Public Health [Internet]. 2025 Jun 23 [cited 2025 Jul 31];13. Available from: https://pubmed.ncbi.nlm.nih.gov/40626156/

75. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature [Internet]. 2015 May 27 [cited 2025 Jul 21];521(7553):436–44. Available from: https://www.nature.com/articles/nature14539

76. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput Sci [Internet].

2021 Nov 1 [cited 2025 Jul 31];2(6):420. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8372231/

77. Debs P, Fayad LM. The promise and limitations of artificial intelligence in musculoskeletal imaging. Frontiers in Radiology [Internet]. 2023 [cited 2025 Jul 21];3. Available from: https://pubmed.ncbi.nlm.nih.gov/37609456/

78. Richie RC. Basics of Artificial Intelligence (AI) Modeling. J Insur Med [Internet]. 2024 Jul 1 [cited 2025 Mar 30];51(1):35–40. Available from: https://dx.doi.org/10.17849/insm-51-1-35-40.1

79. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. 1989 Dec;1(4):541–51.

80. Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging. Radiol Artif Intell [Internet]. 2023 Jul 1 [cited 2025 Jul 31];5(4):e220232. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10388213/

81. Burlina P, Joshi N, Paul W, Pacheco KD, Bressler NM. Addressing Artificial Intelligence Bias in Retinal Diagnostics. Transl Vis Sci Technol [Internet]. 2021 Feb 1 [cited 2025 Jul 31];10(2):13. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7884292/

82. Achour N, Zapata T, Saleh Y, Pierscionek B, Azzopardi-Muscat N, Novillo-Ortiz D, et al. The role of AI in mitigating the impact of radiologist shortages: a systematised review. Health Technol (Berl) [Internet]. 2025 May 1 [cited 2025 Jul 31];15(3):489. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC12085355/

83. Vakalopoulou M, Christodoulidis S, Burgos N, Colliot O, Lepetit V. Deep Learning: Basics and Convolutional Neural Networks (CNNs). Neuromethods [Internet]. 2023 Jul 23 [cited 2025 Mar 30];197:77–115. Available from: https://www.ncbi.nlm.nih.gov/books/NBK597497/

84. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. Sci Robot [Internet]. 2014 Jun 10 [cited 2025 Jul 31];3(January):2672–80. Available from: https://arxiv.org/pdf/1406.2661

85. Wiestler B, Menze B. Deep learning for medical image analysis: a brief introduction. Neurooncol Adv [Internet]. 2021 Dec 1 [cited 2025 Mar 30];2(Suppl 4):iv35. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7829473/

86. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [Internet]. 2013 Nov 12 [cited 2025 Jul 31];8689 LNCS(PART 1):818–33. Available from: https://arxiv.org/pdf/1311.2901

87. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for Simplicity: The All Convolutional Net. 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings [Internet]. 2014 Dec 21 [cited 2025 Aug 28]; Available from: https://arxiv.org/pdf/1412.6806

88. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [Internet]. 2015 Dec 14 [cited 2025 Aug 28];2016-December:2921–9. Available from: https://arxiv.org/pdf/1512.04150

89. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis [Internet]. 2016 Oct 7 [cited 2025 Aug 28];128(2):336–59. Available from: https://arxiv.org/abs/1610.02391v4

90. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018. 2018 May 3;2018-January:839–47.

91. Offiah AC. Current and emerging artificial intelligence applications for pediatric musculoskeletal radiology. Pediatr Radiol [Internet]. 2022 Oct 1 [cited 2025 Jul 21];52(11):2149–58. Available from: https://pubmed.ncbi.nlm.nih.gov/34272573/

92. Vogrin M, Trojner T, Kelc R. Artificial intelligence in musculoskeletal oncological radiology. Radiol Oncol. 2020 Nov 10;55(1):1–6.

93. Hinterwimmer F, Consalvo S, Neumann J, Rueckert D, von Eisenhart-Rothe R, Burgkart R. Applications of machine learning for imaging-driven diagnosis of musculoskeletal malignancies—a scoping review. Eur Radiol. 2022 Oct 1;32(10):7173–84.

94. Zhou X, Wang H, Feng C, Xu R, He Y, Li L, et al. Emerging Applications of Deep Learning in Bone Tumors: Current Advances and Challenges. Front Oncol [Internet]. 2022 Jul 19 [cited 2025 Jul 20];12:908873. Available from: www.frontiersin.org

95. Health AI Register [Internet]. [cited 2025 Aug 28]. Available from: https://healthairegister.com/radiology/products

96. Straus Takahashi M, Donnelly LF, Siala S. Artificial intelligence: a primer for pediatric radiologists. Pediatr Radiol. 2024 Dec 1;54(13):2127–42.

97. He Y, Pan I, Bao B, Halsey K, Chang M, Liu H, et al. Deep learning-based classification of primary bone tumors on radiographs: A preliminary study. EBioMedicine [Internet]. 2020 Dec 1 [cited 2025 Mar 30];62. Available from: https://www.thelancet.com/action/showFullText?pii=S2352396420304977

98. Liu R, Pan D, Xu Y, Zeng H, He Z, Lin J, et al. A deep learning–machine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors. Eur Radiol. 2022 Feb 1;32(2):1371–83.

99. Consalvo S, Hinterwimmer F, Neumann J, Steinborn M, Salzmann M, Seidl F, et al. Two-Phase Deep Learning Algorithm for Detection and Differentiation of Ewing Sarcoma and Acute Osteomyelitis in Paediatric Radiographs. Anticancer Res [Internet]. 2022 Sep 1 [cited 2025 Jul

20];42(9):4371–80. Available from: https://pubmed.ncbi.nlm.nih.gov/36039445/

100. Hinterwimmer F, Serena RS, Wilhelm N, Breden S, Consalvo S, Seidl F, et al. Recommender-based bone tumour classification with radiographs—a link to the past. Eur Radiol. 2024 Oct 1;34(10):6629–38.

101. Hinterwimmer F, Guenther M, Consalvo S, Neumann J, Gersing A, Woertler K, et al. Impact of metadata in multimodal classification of bone tumours. BMC Musculoskelet Disord [Internet]. 2024 Dec 1 [cited 2025 Jul 21];25(1). Available from: https://pubmed.ncbi.nlm.nih.gov/39427131/

102. Xie Z, Zhao H, Song L, Ye Q, Zhong L, Li S, et al. A radiograph-based deep learning model improves radiologists' performance for classification of histological types of primary bone tumors: A multicenter study. Eur J Radiol [Internet]. 2024 Jul 1 [cited 2025 Mar 30];176. Available from: https://pubmed.ncbi.nlm.nih.gov/38733705/

103. Aydin Şimşek Ş, Aydin A, Say F, Cengiz T, Özcan C, Öztürk M, et al. Enhanced enchondroma detection from x-ray images using deep learning: A step towards accurate and cost-effective diagnosis. Journal of Orthopaedic Research [Internet]. 2024 Dec 1 [cited 2025 Jul 21];42(12):2826–34. Available from: https://pubmed.ncbi.nlm.nih.gov/39007705/

104. Do NT, Jung ST, Yang HJ, Kim SH. Multi-Level Seg-Unet Model with Global and Patch-Based X-ray Images for Knee Bone Tumor Detection. Diagnostics 2021, Vol 11, Page 691 [Internet]. 2021 Apr 13 [cited 2025 Jul 28];11(4):691. Available from: https://www.mdpi.com/2075-4418/11/4/691/htm

105. von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Gassert FG, Foreman SC, et al. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. Radiology [Internet]. 2021 Nov 1 [cited 2025 Jul 18];301(2):398–406. Available from: /doi/pdf/10.1148/radiol.2021204531

106. von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Jung M, Jungmann PM, et al. 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 1;32(9):6247–57.

107. Park CW, Oh SJ, Kim KS, Jang MC, Kim IS, Lee YK, et al. Artificial intelligence-based classification of bone tumors in the proximal femur on plain radiographs: System development and validation. PLoS One. 2022 Feb 1;17(2 February).

108. Breden S, Hinterwimmer F, Consalvo S, Neumann J, Knebel C, von Eisenhart-Rothe R, et al. Deep Learning-Based Detection of Bone Tumors around the Knee in X-rays of Children. Journal of Clinical Medicine 2023, Vol 12, Page 5960 [Internet]. 2023 Sep 14 [cited 2025 Jul 20];12(18):5960. Available from: https://www.mdpi.com/2077-0383/12/18/5960/htm

109. Li J, Li S, Li X, Miao S, Dong C, Gao C, et al. Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model. Eur Radiol [Internet]. 2023 Jun 1 [cited 2025 Jul 28];33(6):4237–48. Available from: https://pubmed.ncbi.nlm.nih.gov/36449060/

110. Guo C, Chen Y, Li J. Radiographic imaging and diagnosis of spinal bone tumors: AlexNet and ResNet for the classification of tumor malignancy. J Bone Oncol [Internet]. 2024 Oct 1 [cited 2025 Jul 28];48:100629. Available from: https://www.sciencedirect.com/science/article/pii/S221213742400109X?via%3Dihub

111. Wang H, He Y, Wan L, Li C, Li Z, Li Z, et al. Deep learning models in classifying primary bone tumors and bone infections based on radiographs. npj Precision Oncology 2025 9:1 [Internet]. 2025 Mar 13 [cited 2025 Mar 30];9(1):1–13. Available from: https://www.nature.com/articles/s41698-025-00855-3

112. Technology NI of S and. SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions. 2015 Aug 4 [cited 2025 Mar 30]; Available from: https://csrc.nist.gov/pubs/fips/202/final

113. Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 36th International Conference on Machine Learning, ICML 2019 [Internet]. 2019 May 28 [cited 2025 Jul 30];2019-June:10691–700. Available from: https://arxiv.org/pdf/1905.11946

114. Cullerne Bown W. Sensitivity and Specificity versus Precision and Recall, and Related Dilemmas. J Classif [Internet]. 2024 Jul 1 [cited 2025 Mar 30];41(2):402–26. Available from: https://link.springer.com/article/10.1007/s00357-024-09478-y

115. Dinga R, Penninx BWJH, Veltman DJ, Schmaal L, Marquand AF. Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines. bioRxiv [Internet]. 2019 Aug 22 [cited 2025 Mar 30];743138. Available from: https://www.biorxiv.org/content/10.1101/743138v1

The following tool was used to optimize the language of the text:

- DeepL Translator
- DeepL SE
- 09.09.2025
- https://www.deepl.com/de/translator