Thesis

THE ROLE OF ARTIFICIAL INTELLIGENCE IN DETECTING PEDIATRIC ELBOW FRACTURES

submitted by
Kyrillos Hakeem Boules

in partial fulfilment of the requirements for the degree of

Doktor der gesamten Heilkunde
(Dr. med. univ.)

at the
Medical University of Graz

executed at the
Division of Pediatric Radiology
Department of Radiology

under the supervision of
Sebastian Tschauner, MD PhD
and
Michael Janisch, MD

Graz, 05.11.2024

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, 05.11.2024 *Kyrillos Hakeem Boules m.p.*

I. Acknowledgements

I would like to extend my sincere and heartfelt thanks to my supervisors, Sebastian Tschauner MD, and Michael Janisch MD, for their guidance and support throughout this journey. In particular, I am deeply indebted to Sebastian Tschauner, whose unwavering support, guidance, and encouragement throughout this project have been invaluable. I am truly grateful for the opportunity to work under his mentorship, which has significantly enriched my academic journey and deepened my understanding of this field.

I also want to express my profound appreciation to my family. Their constant love and encouragement have been the backbone of my pursuit of becoming a doctor. They have stood by me through every challenge, celebrated my achievements, and provided me with the motivation to keep striving for my dreams. I am forever thankful for their belief in me and their unwavering support.

I am profoundly thankful as well to my dear friends. In particular, I would like to acknowledge those with whom I shared countless hours of study and camaraderie: Antonios Manoli, Kirolos Gendy, George Abdelnour, Christina Abdelmalak, Ahmed Abaira, Marian Awad, and Jan Urban. The experiences we shared during our studies – the challenges, the laughter, and the many unforgettable moments – transformed this demanding journey into something truly enjoyable. Even the hardest times felt lighter and fleeting with your support, and for that, I am forever grateful.

II. Zusammenfassung auf Deutsch

Forschungszusammenhang

Pädiatrische Ellenbogenfrakturen machen einen bedeutenden Anteil der Traumafälle in der pädiatrischen Notaufnahme aus. Diese Frakturen sind aufgrund der komplexen Anatomie des kindlichen Ellenbogens und der sich entwickelnden Knochenstruktur bei Kindern bekanntermaßen schwer genau zu diagnostizieren. Fehldiagnosen oder verzögerte Diagnosen solcher Frakturen können zu verheerenden Komplikationen führen, einschließlich langfristiger Funktionseinschränkungen oder dauerhafter Behinderungen des betroffenen Kindes. Das derzeit am häufigsten verwendete diagnostische Werkzeug für diese Verletzungen ist die Röntgendurchleuchtung, die trotz ihrer weiten Verbreitung inhärente Einschränkungen in ihrer Sensitivität aufweist. Menschliche Faktoren, wie mangelnde diagnostische Erfahrung bei jungen Ärzten oder Erschöpfung, die auch erfahrene Fachkräfte betreffen kann, können die genaue Beurteilung zusätzlich erschweren. Künstliche Intelligenz (KI) hat bereits gezeigt, dass sie in der Lage ist, in der medizinischen Bildgebung Pathologien präzise zu identifizieren und zu lokalisieren. Mit dieser Forschung zielen wir darauf ab, die Entwicklung von KI-Algorithmen voranzutreiben, die pädiatrische Ellenbogenfrakturen erkennen können, um die diagnostische Genauigkeit zu verbessern und die Behandlungsergebnisse zu optimieren.

Ziele

Diese Arbeit hat folgende Ziele: 1) Zwei verschiedene CNN-Modelle (nämlich EfficientNet und YOLOv8) zu trainieren, um pädiatrische Ellenbogenfrakturen zu erkennen, indem Ellenbogenröntgenbilder von Kindern (welche Frakturen vs. keine Frakturen enthalten) klassifiziert und der „Bereich von Interesse“ (ROI) im Bild markiert wird. 2) Relevante Performancemesswerte (z. B. Precision, Recall, F1-Score, NPV, PR-AUC Kurven und andere) für beide Modelle, EfficientNet und YOLOv8, darzustellen.

Methodik

Wir nutzten einen internen Datensatz, der von der Abteilung für Pädiatrische Radiologie, Institut für Radiologie, Medizinische Universität Graz, zusammengestellt wurde und 17.084 Röntgenbilder von Kinderellenbogen von 6.423 Patienten umfasst. EfficientNet mit seinen Untervarianten B0-B7 und YOLOv8 mit seinen Untervarianten YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv8-l, YOLOv8-x wurden trainiert, um kindliche Ellenbogenfrakturen auf den Bildern zu erkennen und zu lokalisieren, wobei letzteres durch das Markieren des Bereichs von Interesse (ROI) mit Begrenzungsrahmen erfolgte. Nach dem Training beider Modelle haben wir deren Leistung anhand standardisierter KI-Metriken wie Accuracy, Precision, Sensitivität und F1-Score sowie weiterer Metriken bewertet. Diese Metriken wurden dann genutzt, um die Effektivität der EfficientNet-Untervarianten und der YOLOv8-Untervarianten bei der genauen Erkennung und Lokalisierung von kindlichen Ellenbogenfrakturen miteinander zu vergleichen. Für alle getesteten EfficientNet-Untervarianten und YOLOv8-Untervarianten wurden PR-AUC Kurven erstellt.

Ergebnisse

Unter den EfficientNet-Varianten erzielte das Modell B3 die höchsten Werte in Bezug auf Recall (94,41%), F1-Score (92,22%) und Accuracy (88,6%). Das Modell B6 hingegen erreichte die besten Ergebnisse in den Kategorien Precision (91,62%), Spezifität (78,87%), und NPV (91,62%), wodurch es besonders effektiv in der Minimierung von falsch-positiven Diagnosen ist. Die PR-AUC von B0 beträgt 0,94, die von B2 und B4 liegt bei 0,95, die von B1, B3 und B7 bei 0,96 und schließlich die von B5 und B6 bei 0,97.

In der YOLOv8-Familie erreichte YOLOv8-x die höchste Precision (94,4%), F1-Score (91,45%) und AP50-95% (66,12%), während YOLOv8-n und YOLOv8-s die höchsten NPV-Werte (beide jeweils mit 97,64%) aufwiesen. YOLOv8-m, YOLOv8-l und YOLOv8-x haben jeweils eine PR-AUC von 0,954, während YOLOv8-n und YOLOv8-s jeweils eine PR-AUC von 0,951 haben.

Diskussion

Sowohl die EfficientNet- als auch die YOLOv8-Modelle zeigten ein großes Potenzial, die Erkennung von pädiatrischen Ellenbogenfrakturen zu unterstützen, wobei ihre Leistung je nach spezifischen Metriken variierte. Unsere Ergebnisse unterstreichen das Potenzial von KI-Modellen, die radiologische Diagnostik im pädiatrischen Bereich zu verbessern. Allerdings sind weitere Verfeinerungen und klinische Validierungen erforderlich, um ihre Wirksamkeit in realen Anwendungen sicherzustellen.

III. Abstract

Big picture of research context

Pediatric elbow fractures constitute a significant portion of trauma cases seen in pediatric emergency departments. These fractures are notoriously difficult to diagnose accurately due to the complex anatomy of the pediatric elbow and the evolving nature of bone development in children. Misdiagnosis or delayed diagnosis of such fractures can lead to devastating complications, including long-term functional impairment or permanent disability in the affected child. Currently, the standard diagnostic tool for these injuries remains radiography, which, while widely used, suffers from inherent limitations in sensitivity. Human factors, such as diagnostic inexperience in junior clinicians or fatigue affecting even seasoned professionals, can further complicate accurate assessment. Artificial intelligence (AI) has already demonstrated the potential to assist in medical imaging by accurately identifying and localizing pathologies in various radiological studies. With this research, we aim to advance the field by developing AI algorithms capable of detecting pediatric elbow fractures, contributing to improved diagnostic accuracy and patient outcomes.

Aims

This thesis 1) aims to train two different CNN models (namely EfficientNet and YOLOv8) to detect pediatric elbow fractures by classifying pediatric elbow radiographs (fracture vs no fracture) and framing the region of interest (ROI), 2) aims to present relevant performance metrics (e.g. precision, recall, F1-score, NPV, PR-AUC curves among others) for both EfficientNet and YOLOv8 models.

Methodology

We utilized an internal dataset curated by the Division of Pediatric Radiology, Department of Radiology, Medical University of Graz, which comprises 17,084 pediatric elbow X-ray images from 6,423 patients. EfficientNet with its sub-variants B0-B7 and YOLOv8 with its sub-variants YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv8-l, YOLOv8-x were trained to detect and localize pediatric elbow fractures

within the images, the latter by marking the region of interest (ROI) with bounding boxes. After training both models, we evaluated their performance using standard AI metrics, such as accuracy, precision, recall, and F1-score among others. These metrics were then used to compare the effectiveness of the EfficientNet's sub-variants and YOLOv8's sub-variants with one another in accurately detecting and localizing pediatric elbow fractures. PR-AUC curves were created for all tested EfficientNet variants and YOLOv8 variants.

Results

Among the EfficientNet variants, the B3 model achieved the highest recall (94.41%), F1-score (92.22%), and accuracy (88.6%). The B6 model, on the other hand, excelled in precision (91.62%), specificity (78.87%), and NPV (91.62%), making it particularly effective at minimizing false positives. B0 achieved a PR-AUC of 0.94, B2 and B4 have a PR-AUC of 0.95, B1, B3 and B7 have a PR-AUC of 0.96, and finally B5 and B6 have a PR-AUC of 0.97.

In the YOLOv8 family, YOLOv8-x achieved the highest precision (94.4%), F1-score (91.45%), and AP50-95% (66.12%), while YOLOv8-n and YOLOv8-s tied for the highest NPV scores (97.64%). YOLOv8-m, YOLOv8-l, YOLOv8-x have each a PR-AUC of 0.954 whereas YOLOv8-n, and YOLOv8-s each have a PR-AUC of 0.951.

Discussion

Both EfficientNet and YOLOv8 models demonstrated strong potential for assisting in the detection of pediatric elbow fractures, though their performance varied depending on specific metrics. Our findings underscore the potential for AI models to enhance radiographic diagnosis in pediatric settings, though further refinement and clinical validation are necessary to ensure their effectiveness in real-world applications.

Table of Contents

IV. Abbreviations and their meaning

a.p.	anterior-posterior (view)
AI	artificial intelligence
ALARA	as low as reasonably achievable
AP50%	average precision at a confidence threshold of 50%
AP50-95%	average precision at a range of IoU thresholds from 50% to 95%
BMP	bitmap (image file)
CNN	convolutional neural network
CONV	convolutional (layer)
CSP	cross stage partial
CT	computed tomography
CV	computer vision
CVS	computer vision systems
DICOM	digital imaging and communications in medicine
DL	deep learning
FC	fully connected (layer)
FN	false negative
FP	false positive
FRACTURE	fast field echo resembling a CT using restricted echo-spacing
ID	identification number
IoU	intersection over union
JPG	joint photographic experts group
lat.	lateral

MeSH	medical subject headings
mo	months
MRI	magnetic resonance imaging
NPV	negative predictive value
PACS	picture archiving and communication system
PANet	path aggregation network
PNG	portable network graphics
POOL	pooling (layer)
PPV	positive predictive value
PR-AUC	precision-recall area under the curve
ResNet	residual neural network
ROI	region of interest
TIF	tag image file (format)
TN	true negative
TNR	true negative rate
TP	true positive
TPR	true positive rate
US	ultrasound
wks	weeks
XAI	explainable artificial intelligence
XAOM	x-ray alignment and orientation method
YOLOv8	you-only-look-once version 8
yrs	years

VI. Table of figures

VII. List of tables

1. Introduction to the thesis

1.1. Artificial intelligence

Artificial Intelligence (AI) as a field originated in the 1950s. John McCarthy, who is perceived as the godfather of modern AI, was the first one who coined the expression “Artificial Intelligence” (1). AI can be broadly defined “*as the capability of a machine to imitate human behavior*” (2). The main selling point of using AI in today’s world is its capability of analysing data, using a sub-area of AI named “deep learning” (DL). AI algorithms are able to learn in a quick fashion, and in doing so recognizing trends, correlations, anomalies which could be easily missed by human analysis (3,4). AI is currently used in many fields of medicine such as radiology, pathology, oncology, cardiology, gastroenterology, ophthalmology, and even surgery (2). In the field of diagnostic radiology, AI has proven to be helpful in improving diagnostic accuracy and efficiency in most imaging modalities such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound (US), so that it enables professional healthcare providers to detect pathologies early and provide personalized therapy plans. In a lot of cases so-called convolutional neural networks (CNNs), which are a subset of machine learning algorithms, are applied to develop image recognizing AI, which are also known as “computer vision solutions” (CVS). Computer vision solutions, when trained on big sums of datasets, are afterwords able to classify images, detect objects, perform image segmentation into various different regions as well as produce captions for images (5).

1.2. Why we should care about implementing AI in diagnostic imaging

The experience level of diagnosticians plays a crucial role in finding the right diagnosis and thereby an adequate therapeutic plan, as for example Manning D. and colleagues have shown (6). AI has successfully proven to be able to compensate diagnostician’s inexperience, for beginners as well as seniors (7); it might as well serve as a reliable “second opinion tool” in the far future (8). Diagnosticians fatigue is another serious limitation that can adversely affect patient’s outcomes (9). AI does not get exhausted; it does not fall out neither intendedly (e.g.

go on vacation) nor casually (e.g. be on sick leave). It does not have motivational fluctuations or mood swings triggered by stress, lack of time or personal circumstances. It is rather operating very consistently. The number of qualified health care providers decreases (10), meanwhile diagnosis requests keep increasing (11). A big problem of a lot of AI models is, that we as humans often do not know how or why its algorithms come to a certain result/prediction. For these types of AI the term “black box AI” is used to describe the fact, that all its decision making operations are hidden from the user (12). A relatively new emerging subfield of AI called explainable AI (XAI) tackles this problem by helping explain how the model arrived at certain outcomes/predictions, via different methods. In the future XAI models will be able to communicate what the AI algorithm is doing using written humanly understandable natural language. Soon we as humans might be able to discover new correlations that we were not aware of which were taught to us by XAI (13).

Furthermore, bigger advancements in AI could conduce to radiation protection by making the need of a second X-ray view obsolete (14). Based on study conclusions like the latter one, we can logically extend it to suggest that radiographs of the uninjured contralateral side, often taken for comparison, could also become unnecessary. If AI can accurately diagnose fractures or abnormalities from a single view, eliminating the need for multiple projections (e.g. anteroposterior and lateral views), it stands to reason that comparative imaging of the healthy side might no longer be required either.

Opposite side comparison X-rays are sometimes performed due to uncertainty, to find out whether the findings in the x-ray of the injured side are physiological or pathological. Elbow joint injuries are counted among the predestined conditions to undergo “opposite side radiography” because of the joint’s multiple ossification centers that mimic fractures and consecutively add uncertainty to diagnosticians (15).

With AI's growing ability to identify and analyse subtle differences in a single view, we could avoid additional imaging not just for multiple angles but also for

comparison with healthy tissues, further contributing to minimizing patient radiation exposure.

Finally, AI can help fight healthcare disparities, but only if its usage is not limited to locations which benefit from extensive healthcare availability already (16).

1.3. PubMed search: AI detection of elbow fractures in children

When searching PubMed, using medical subject headings (MeSH), for articles that deal with the utilization of artificial intelligence to detect specifically elbow fractures in children, currently (as of the 30th of July 2024) only 2 papers are shown. In my search I used the terms ("Artificial Intelligence"[Mesh]) AND "Elbow Fractures"[Mesh]), 3 papers were shown to me from which only two were related to children. The first one was published by Dupuis M. and colleagues, who dealt with the *"External validation of an artificial intelligence solution for the detection of elbow fractures and joint effusions in children"* (17). In this study a commercially available DL algorithm for detecting elbow fractures (namely "SmartUrgences") was put to the test, by letting it look through digital radiographs from a real-life cohort of children presenting to a single-centered emergency room. 4 assessments were done, including the capability of the DL algorithm to detect elbow fractures and/or joint effusions. The fracture detecting algorithm accomplished a recall of 92.9%, a precision of 70.8% and an F1-score of 0.804, which is considered to be quite good. However, 8% of fractures were missed by the AI, by the means of classifying them as false negatives. The authors claim that the 8% of cases that were diagnosed incorrectly, is a too high proportion for this AI solution to be a self-sufficient screening tool, especially when considering that those 8% were classified as having no fractures when in reality a fracture was present (17).

The second study was written by Kekatpure A. and colleagues (18), and it aimed to generate a DL based support system to detect distal humerus fractures using elbow radiographs. X-rays of children under the age of 7 were excluded due to the physiological missing of the ossification centers of the capitellum, radial head, internal epicondyle, and the trochlea. 4,931 elbow radiographs (in whether a.p. or lateral view or both) which were cropped to the area of the capitellum were included.

ResNet18 was the DL algorithm architecture used. AI reached a sensitivity of 61.49%, a specificity of 95.89%, a positive predictive value (PPV) of 99.47%, a negative predictive value (NPV) of 16.59% and a consecutive F1-score of 0.76. The false positive rate is 4,1% which is considered to by high by the authors. So, in conclusion this algorithm had a high precision (99.47%), therefore when it classifies a radiograph to “contain a fracture”, it is in 99.47% of cases right. On the other hand, the model is not very accurate when trying to identify the healthy patients as the false positive rate is with 4% relatively high. The high precision and specificity come at the cost of the low sensitivity of the model. The authors conclude that further research and validation are essential to further improve overall sensitivity of the model to be a truly helpful tool in daily clinical medicine (18).

On balance, one can say that the number of models which can flawlessly detect elbow fractures in children is zero at the moment and does not meet the needed standard for self-sufficient daily clinical practice quit yet. On the other hand, the current results show a remarkable potential of using AI to reach this goal in the future. To achieve this though, larger amounts of data must be introduced to such models and further research in the development of more advanced DL algorithm architectures must be made.

1.4. Features of pediatric bones and common pediatric fracture types

The pediatric periosteum is thicker, firmer, and much more vascularized compared to adults. Due to its looser connection with the underlying bone in comparison to adults, it typically remains intact after trauma (19).

Bones of children feature greater elasticity and plasticity than adult bones. These characteristics can lead to a sometimes (partially) reversible plastic deformation of a bone, instead of a manifest "fracture" in the typical sense, if the applied force was relatively low in energy (20,21). If the deformation is not reversible and therefore radiographically detectable but the bone did not crack, the term “bowing fracture” is used. The latter type of fracture together with the so-called “torus fractures” and “greenstick fractures” belong to the group of “incomplete fractures” (21). Torus fractures usually emerge when axial force is exerted along the longitudinal axis of

the bone, leading to trabecular compression which appears as a bulging of the bone shaft's cortex in X-ray imaging. The radius is one of the most common bones to be thereby involved (22). Greenstick fractures on the other hand result into a fracture of the convex side of a bone while the concave side stays intact and/or is bent. The causative force in these cases is commonly applied vertical to the bone or sometimes longitudinal in an angulated fashion (23).

Two additional characteristic features of growing pediatric bones are the presence of epiphysial plates (= growth plates) and ossification centers. A growth plate is a specific transformation zone between the epiphysis and diaphysis of the bone, composed of cartilage tissue that eventually becomes bone tissue, and is essential to the longitudinal growth of a bone. Any disturbance to this specific zone, for instance due to a fracture that involves it, can lead to growth restraint. Approximately 15-20% of juxta-articular fractures in children involve the growth plate. When it comes to injuries of the growth plate a distinction is made between simple epiphysiolysis (= a smooth transverse fracture which spans the whole growth plate) without an accompanying bone fracture, and those which are combined with additional epi- or metaphysis bone fragmentation (see also Salter-Harris-classification) (19).

Ossification centers are tissue zones from which ossification (= bone formation) originates. There are two types of ossification centers, primary and secondary. In the 8^{th} or 9^{th} gestational month all the primary ossification centers are established. These emerge in long bones in the diaphysis and ossify it together with the metaphysis. The epiphysis however does not get ossified yet because it contains the growth plate. Postnatal multiple secondary ossification centers arise at various times, but in a certain chronological order from the (at that point in time) cartilaginous epiphysis and shape it further. Out of the latter two facts "the actual bone age" can be derived. In the end all secondary ossification centers fuse together and the growth plate closes, and the person is fully grown (19).

From an age of 7 until the age of puberty so-called apophyses develop, which are bony processes. They step out of separate ossification centers and are connected

by a cartilaginous bridge to the epiphysis. The fusion of the apophyses with the epiphyses starts after puberty and ends with the 25th year of live (19).

Avulsion fractures can occur wherever soft tissue, like tendons or ligaments, is connected to a bone. These fractures are often caused by a sudden, forceful pull on the soft tissue that exerts tension on the bone, resulting in the bone fragmenting. Additionally, chronic repetitive stress can also lead to a piece of bone being torn away by the attached tissue over time (24). Osteocartilaginous interfaces (as they are present in growth plates as well as when secondary ossification centers or apophyses develop), along with the nearby newly developed bone, are the most vulnerable areas in a growing skeleton. Therefore, avulsion fractures occur much more often in adolescents due to their growing muscle strength and the relative weakness of their osteocartilaginous junctions. In contrast musculotendinous junctions constitute the weakest links in a fully developed adult skeleton. Therefore, on the other hand, musculotendinous injuries tend to occur more commonly in adults due to a relatively greater bone strength in comparison to the strength of the musculotendinous interface (25).

Transverse, oblique, or spiral fractures rupture the periosteum and break the bones continuity. The fracture ends get dislocated. Fracture ends in transverse fractures appear in X-ray imaging as a serrated line in contrast to pathological fractures, where the ends appear straight (19).

Fracture healing proceeds in pediatric patients significantly quicker compared to the adult population (20). A fracture typically heals by formation of a callus, within the meaning of secondary fracture healing. The younger the patient the bigger the callus (19). Since certain axial misalignments self-correct with growing, ensuring precise anatomical realignment is not always essential. However, as mentioned above, fractures involving the growth plate can lead to significant deformities (20). Although only a small proportion of pediatric bone fractures require surgical fixation (26), there is evidence (27) that the number of such interventions has increased over time. Bryson D. J. and colleagues suggest that the reason behind an urge to perform early surgical fixation might be *“driven by the desire for immediate correction of clinical*

deformities or by financial incentives to minimise hospital costs accrued with prolonged admissions during periods of immobilisation." (27).

The variety of potential fracture types and their age-specific manifestations in both clinical examinations and imaging lead to a relatively flat learning curve for diagnosticians (28). Extensive expertise and training are thus essential to maintain a high quality of diagnostic and therapeutic precision when treating injured children (29).

1.5. Diagnostic imaging of pediatric fractures

Missing pediatric fractures can have devastating long-term effects such as long-term disability and increased mortality. Pediatric fractures are the most common reason for medical-legal issues and account for the biggest source of malpractice claims in the emergency department. The main cause for this is the deficit in expertise of junior physicians in child skeletal injuries (30).

Plain radiography still serves as the indispensable modality to diagnose pediatric bone fractures both initially and in follow-up. Usually, radiographic studies are performed in two planes that are orthogonal to each other (most commonly: "anterior-posterior view" and "lateral view"). In cases in which the deformity of the harmed area is clinically visible and the radiograph shows a clear diagnosis in one plane, then the performance of the second plane for additional diagnostic purposes may be useless (31).

The additional usage of ultrasound (US), computed tomography (CT) and magnetic resonance imaging (MRI) can be indicated in cases, where a definitive fracture diagnosis through X-ray imaging only, remains uncertain (32). However there are cases where all fractures diagnosis can be missed in all of these modalities (33).

CT plays an important role in the diagnosing of complex fractures (34). Some of its most valuable areas of diagnostic usage are pediatric orbital fractures, skull fractures and thoracolumbar compression fractures (35–37). Also its crucial role in preoperative planning of intraarticular fractures remains undisputed (38). Since extended CT radiation exposure may increase the risk of radiation-induced cancer, especially in children, the abidance of the "ALARA"-principle (= as low as reasonably

achievable) is all the more important in this age group (39). Hence the quantity of useful indication to use CT in pediatric traumatology is strictly limited (40).

There is a lot of evidence that MRI is another useful tool in detecting fractures, even in the cases that are not visible in radiographs (41). Traditionally CT is the gold standard imaging to detect bone fractures especially when surgeries need to be planned in the future. However recent studies have shown that certain MRI techniques like e.g. “FRACTURE” (= **f**ast field echo **r**esembling **a** **CT** **u**sing **r**estricted **e**cho-spacing) can provide similar cortical and trabecular bone contrast to CT scans, without radiation exposure. The latter advantage and the similar outcome of diagnostic accuracy could potentially result into a replacement of the performance of preoperative CT scans by techniques like this in the future in certain conditions (42,43). The biggest selling point to use MRI instead of other modalities is its excellent ability to visualize soft tissue like e.g. bone marrow or ligaments, making it the mostly used modality to diagnose injuries of the same such as ligament ruptures or bone marrow edema (44).

Ultrasound has been shown to be a sufficient modality in diagnosing various types of pediatric fractures (45). Some studies report that, certain types of pediatric fractures should be initially assessed by ultrasound only, e.g. skull, non-displaced distal forearm and clavicle fractures (46). Ultrasound may as well be used in the early detection of impaired fracture healing (47). Furthermore ultrasound can also be used therapeutically to help fracture healing especially in cases of malunion (48).

1.6. Closer look: Pediatric elbow and its fractures

Pediatric elbow fractures are the second most common pediatric fractures (19) and make from 15 up to 40% of all observed pediatric fractures. Despite being relatively common, these injuries can be among the most difficult to detect (49). Supracondylar humerus fractures are the most frequent seen elbow fracture in children and account for 10% of all pediatric fractures. Transcondylar fractures represent 14% of pediatric elbow fractures and secure thereby the second place (19).

1.7. Difficulty in diagnosing elbow fractures

Elbow joint fractures are difficult to detect because of the invisibility of radiolucent cartilage and, ossification centers that appear at various times. Ossification centers can be easily misinterpreted as bone fragments or loose joint bodies, thus as a fracture. Vice versa actual fractures can be easily misconceived for ossification centers and remain undetected (19,50). *"No other joint is more frequently radiographed contralaterally as the result of confusion or uncertainty about the appearance of its normal structures"* (51).

1.7.1. Maturation sequence of the elbow's ossification centers

The multiple secondary ossification centers that build the elbow arise in certain spaces during lifetime in a particular order. In girls this maturation sequence tends to begin a little earlier compared to boys. One can memorize the correct order by using the mnemonic "CRITOE". Reading it from left to right, each letter stands for an anatomical structure from where the next ossification center in the maturation sequence is going to emerge (52). Figure 1 and Figure 2 show the regions where ossification centers of the elbow emerge. Table 1 contains the age of appearance and the age of fusion of the elbow's ossification centers:

	Age of appearance range (average)		**Fusion age**
	Boys	Girls	
C apitellum humeri	6 wks - 11 mo (5 mo)	1-7 mo (4 mo)	12 yrs
R adial head	3 – 6 yrs (5 yrs)	3 – 6 yrs (4 yrs)	15 yrs
I nternal (= medial) epicondyle	5 – 7 yrs (7 yrs)	3 – 6 yrs (5 yrs)	17 yrs
T rochlea	8 – 9 yrs (9 yrs)	7 – 9 yrs (8 yrs)	12 yrs
O lecranon	8 – 10 yrs (10 yrs)	8 yrs	15 yrs

E xternal (= lateral) epicondyle	11 – 12 yrs (12 yrs)	11 yrs	12 yrs

Table 1 Age of appearance and fusion of elbow's ossification centers, modified after (52).

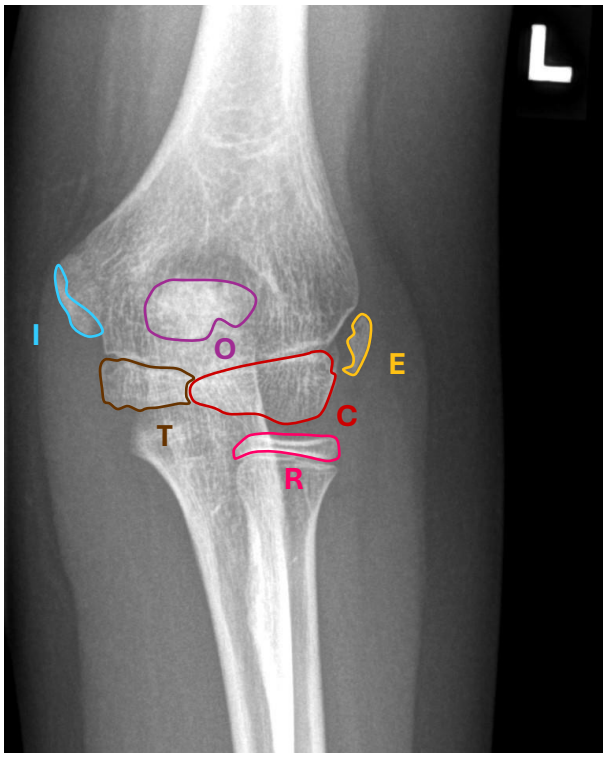

Figure 1 Normal radiographs of the elbow in a 11-year-old child. The ossification centers are annotated. (a.p. view).

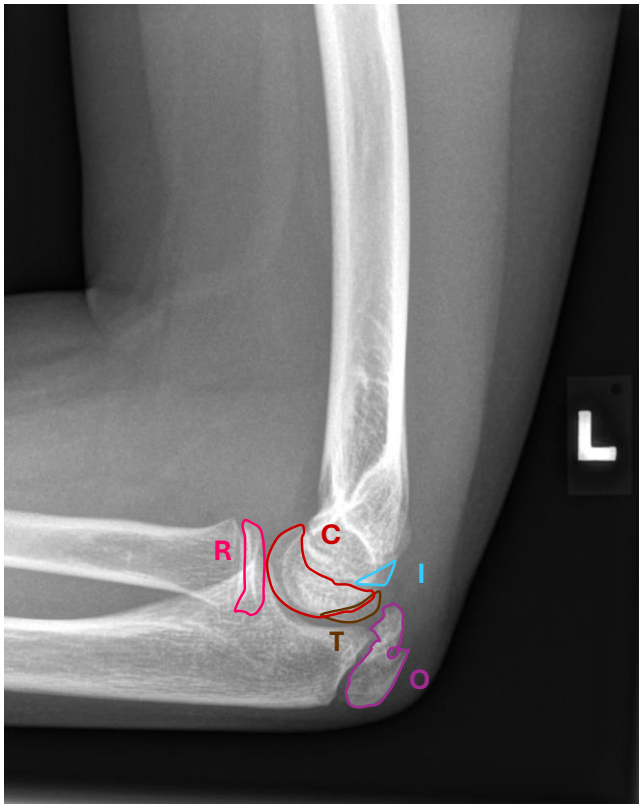

Figure 2 Normal radiographs of the elbow in a 11-year-old child. The ossification centers are annotated. (lat. view).

1.7.2. Imaging elbow fractures

Generally, the anterior-posterior (a.p.) together with the lateral (lat.) X-ray view are enough to be able to diagnose an elbow fracture. The path of the rays in the lateral view should be always radioulnar (19). However, some studies suggest, that at certain emergency departments, more than a quarter of all elbow fractures can be missed by diagnosticians in radiographic studies (53).

1.7.3. Measurements and radiographic signs to help detect elbow fractures

1.7.3.1. Findings on the lateral view

The elbow joint is equipped with 3 fat pads, an anterior and posterior one, and the third one lies parallel to the anterior radius and is called supinator fat pad. The first two fat pads are located within the articular capsule but outside the synovial fluid. The anterior fat pad lies in the radial and coronoid fossae of the anterior humerus, whereas the posterior fat pad lies within the olecranon fossa on the posterior aspect of the humerus. In a normal lateral view X-ray, merely the anterior and supinator fad

pads are seen, while the posterior one is invisible due to its depression in the olecranon fossa. Thereby the anterior pad fat looks like a radiolucent tear drop and the supinator fat pad is portrait as a very thin sausage-shaped radiolucent area. After an elbow injury occurs, intraarticular joint effusions cause a displacement of the fat pads or modify their shape (54). Cave: In the event of rupture of the joint capsule and consecutive elbow dislocation, fat pad signs could be invisible due to decompression of the joint effusion (55).

Sail sign

In case of a reshaping of the anterior fat pad into a convex triangular sail-like form an injury to the elbow is very likely (49).

Posterior fat pad sign

Approximately 70% of pediatric elbow fractures hide behind present posterior fat pads. Since the olecranon fossa takes the posterior fat pad fully in, only moderate joint effusions are necessary to cause their pathological displacement. Thus, a positive posterior fat pad sign is associated with a high specificity to elbow injuries especially fractures (55).

Ventrally shifted supinator fat pad

There is a fifty-fifty chance that a more ventrally shifted supinator fat pad indicates a radial neck fracture (55).

Roger-Line: Anterior Humeral Line (AHL)

Normally a straight line drawn along the anterior aspect of the humeral cortex should intersect with the middle third of the capitellar ossification center. In children younger than the age of 4, this line intersects regularly with the border zone between the anterior to middle third of the capitellum and constitute a normal finding. If the intersection point is positioned in the furthest anterior third of the capitellum or missing it by lying in front of it, a displaced extension supracondylar fracture should be suspected. Comes the intersection line to lie in the posterior third of the capitellum or misses it by lying behind it, a slightly dislocated flexion fracture can be causative (49).

Radiocapitellar Line

A straight line drawn through the proximal radial shaft should pass through the center of the capitellum humeri in all radiological projections regardless of the position of the elbow or forearm (19,49). If this line does not intersect the center of the capitellum humeri, a radial head dislocation is likely (19). In children younger than the age of 2, this rule is only in half of cases consistently true. Some authors therefore suggest aligning the line in these children between the neck of the radius instead of its shaft. The line should transect the capitellum, if not an injury might be causative. The radiocapitellar line is the most practical to help detecting Monteggia fractures, which is essentially a fracture of the ulna's shaft combined with a radial head dislocation (49).

1.7.3.2. Findings on the a.p. view

Humeroulnar angle

This landmark is investigated by measuring the angle between the axis of the humeral shaft and the axis of the ulnar shaft. Both lines should equally bisect their respective bones in a valgus posture. A normal reference rage would be from 5° - 15° for this angle. An angle less than 5° or an elbow in a clinically varus posture, could indicate either a lateral condyle fractures or a supracondylar humerus fractures (49).

Baumann's angle: Humerocaptitellar angle

Baumann's angle is measured by investigating the angle which is formed by the axis of the humeral shaft and a line through the growth plate of the lateral epicondyle. A normal reference rage would be from 64° - 81° for this angle. Supracondylar fractures would widen this angle (49).

1.8. Trauma imaging and computer vision solutions

The biggest selling point for the usage of artificial intelligence in medicine as well as in other fields, is its ability to analyze large sums of data in a fairly short time (56). When it comes to computer vision (CV) solutions, which allow computers to recognize objects in digital images and videos, usually deep learning (DL) methods that are built on so called convolutional neuronal networks (CNNs) are used (57). The most important capabilities of computer vision algorithms are

- **image classification:** Assign a label or category to an entire image based on its content.
- **object detection:** Identify and locate specific objects within an image or video frame, often with bounding boxes.
- **object localization:** Determine the precise location of objects within an image, often represented by bounding boxes or other geometric shapes.
- **image segmentation:** Partition an image into multiple segments or regions to simplify or change the representation of an image. This includes:
    - **semantic segmentation:** Assign a label to every pixel in an image.
    - **instance segmentation:** Detect and delineate each distinct object of interest in an image (58).

Applying those terms in conjunction with osseus trauma their definition would be as the follows.

- **Fracture classification** involves using an algorithm to determine the likelihood that an input image shows a fracture or differentiate between fracture types one another (59).
- **Fracture detection** outputs a region of interest (ROI) which is assumed to include a fracture, often marked by a bounding box (60).
- **Fracture segmentation** is usually more accurate than fracture detection since it predicts the likelihood of individual pixels being part of a fracture. This method offers a detailed outline of the fracture area within the X-ray image, resulting in a more precise and localized identification. In contrast, detection typically only signals whether fractures are present or not (61).

The underlying dataset is the foundation for any computer vision (CV) model. It is a standard practice to divide CV datasets into three distinct, non-overlapping parts:

1. **“Training set”** which is the largest sub-set of images. It contains the actual training data.
2. **“Validation set”** which is a sub-set of the data used to monitor and validate the learning progress during training.
3. The **“Test set”** comprises images that were not shown to the algorithm during the training process. Ideally, this set is supplied by an external

institution, allowing for an unbiased evaluation of the model's generalization capability (62).

Class-imbalance (as in disparities of samples per group, age, or visible features) lead on to selection biases, and in doing so can have adverse effects in the model's overall performance. Well balanced training sets help to prevent CV models from suggesting false conclusions (63).

2. Hypothesis

Digital X-rays of pediatric fractures may be misinterpreted, even though the required diagnostic information is at least partially, if not fully, present within the image pixels. This could result in incorrect or postponed treatment, heightened discomfort, prolonged recovery periods, or even long-term impairment. We propose that AI may outperform human specialists in identifying pertinent diagnostic information, determining the need for further imaging, and anticipating complications in the initial digital radiographs of injured children.

3. Aims

3.1. Training two different CNN models to detect and localize pediatric elbow fractures

The primary objective of this study is to train and evaluate the effectiveness of two different convolutional neural network (CNN) models—YOLOv8 and EfficientNet—in detecting pediatric elbow fractures from radiographs. Specifically, we aim to compare the performance of various YOLOv8 sub-variants (YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv8-l, YOLOv8-x) and EfficientNet sub-variants (B0-B7) against one another to determine which model achieves the highest performance in identifying both fractured and non-fractured elbow images. A crucial aspect of our research is to extract and present relevant performance metrics for both YOLOv8 and EfficientNet models. By quantifying key performance indicators such as sensitivity, specificity, precision, and overall accuracy, we can facilitate a comprehensive comparison between the models and existing CNN frameworks.

4. Materials and methods

The study protocol (No. EK 31-108 ex 18/19) was approved by the ethics committee of the Medical University of Graz (IRB00002556). The ethic committee did not require informed patient or legal representative consent, because of the retrospective nature of this study's data analysis. All study related methods were performed pursuant to the Declaration of Helsinki and the relevant regulations and guidelines. Additionally, it is possible to make the dataset available if reasonably requested.

4.1. Dataset

The dataset utilized in the research was provided by the Division of Pediatric Radiology at the Department of Radiology, Medical University of Graz, Austria, and consisted of X-rays performed between 2008 and 2018. There the local pediatric elbow X-ray dataset was searched for studies that included traumatic (without orthopedic) elbow fractures within that time period. 17,084 pediatric elbow radiographs, with or without elbow fractures, from 6,423 unique pediatric patients were involved in this study. The radiographs were labeled by various radiologists and students through the online "Computer Vision Annotation Tool (CVAT)". The labelers marked areas containing elbow fractures by drawing bounding boxes around them. Although three board certified pediatric radiologist approved all annotations, the occurrence of inevitable inaccuracies, discrepancies, and errors in labeling due to the restricted diagnostic sensitivity of X-ray studies must be taken into consideration (64). These radiographically present fractures were not confirmed by any other radiologic modality such as CT or MRI. Table 2 gives an outline of the dataset's characteristics. 3,463 of the participating patients were males, 2954 were females and 6 whose gender was unknown. The youngest patients were newborns and the oldest 18,8 years old. The average age was 7.31 ± 3.59 based on all genders, and 7.46 ± 3.62, 7.14 ± 3.56 and finally 6.21 ± 1.40 for the male, female and unknown gender participants respectively. Out of the 17084 X-rays, 5,420 did not contain fractures, 11,248 contained one fracture and 416 contained two or more fractures. In the Appendix, Figure 3 displays a mosaic illustrating the diversity of X-ray images in the dataset.

The dataset was divided into three randomly sampled disjoint subsets:

- “Training dataset” which contained 14096 (82,50%) images.
- “Validation dataset” which contained 2488 (14,56%) images.
- “Test dataset” which contained 500 (2,93%) images.

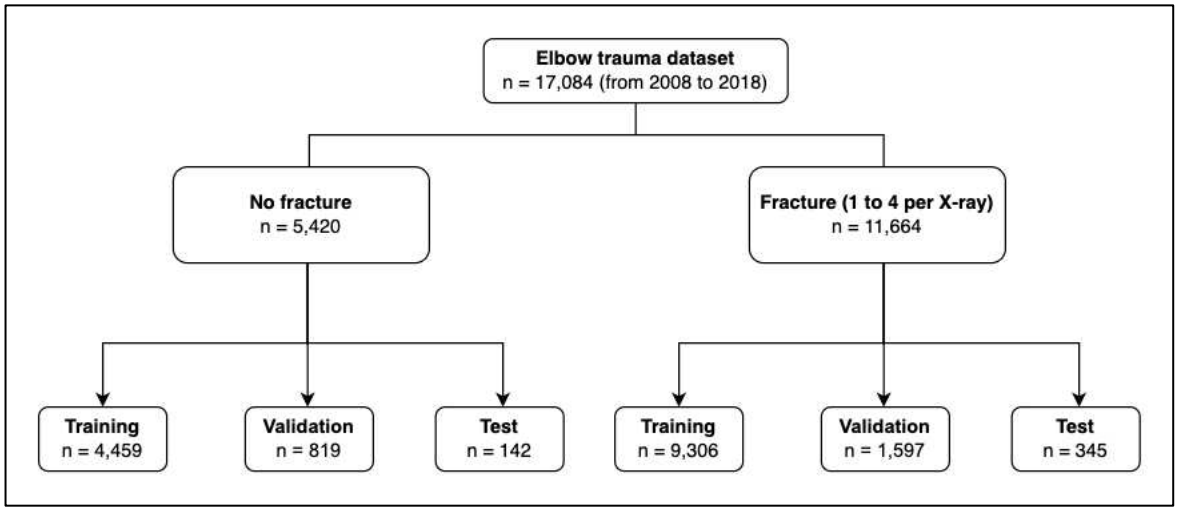

Figure 3 Diversity of X-ray images in the dataset.

		Training set	Validation set	Test set	total (n = 17,084)
age	mean (yrs)	7.31	7.33	7.29	-
	SD (yrs)	3.59	3.61	3.62	-
	min (yrs)	0	0.8	1.2	-
	max (yrs)	18.8	18.0	18.8	-
gender	male	7,554	1,343	260	9,157
	female	6,526	1,140	240	7,906
	unknown	16	5	0	21

laterality	left	4,422	809	148	5,379
	right	3,137	548	125	3,810
	unknown	6,537	1,131	227	7,895
projection	a.p.	3,724	674	128	4,526
	lateral	4,420	782	162	5,364
	oblique	32	4	1	37
	unknown	5,920	1,028	209	7,157
fracture	none	4,459	819	142	5,420
	one	9,306	1,597	345	11,248
	two	322	70	13	405
	three	8	1	0	9
	four	1	1	0	2

Table 2 Outline of the dataset's characteristics.

The dataset was assembled by combining three key elements: pediatric elbow X-rays (image data), associated report texts (natural language descriptions), and expert annotations. Expert annotations included bounding boxes, lines, polygons, and image tags.

4.2. Image query and preprocessing

All 17,084 pediatric elbow X-rays had been obtained as Digital Imaging and Communications in Medicine (DICOM) images from the local Picture Archiving and Communication System (PACS). DICOM is the standard format for medical images, allowing for the long-term storage and seamless exchange of medical images between various imaging modalities and healthcare institutions (65,66).

4.3. Patient de-identification

To de-identify and convert the images in batches, a custom-coded Python script was utilized. PNGs don't support storing metadata as a standard feature. To preserve necessary information for processing and analysis, we generated filenames based on DICOM header data. Identifiers unique to each patients were created from DICOM header information using two algorithms, namely "Blake2b"(67) and “SHA-3-256” (68). The resulting hash remains consistent for each individual patient. The second part of the filename was composed of the acquisition time’s Unix timestamp, minus a 9-digit hash derived from the “Patient ID,” which maintained the original study intervals while irreversibly obscuring the actual examination times. The next section of the filename was formed by the series number and image number (zero-padded to two digits), and is followed by a region token, the laterality und projection, and ending with the patient’s sex and age (rounded to one decimal place). In summary an example file name could be as shown in Figure 4. The exact coding steps remain with our institution to minimize risks of re-identification:

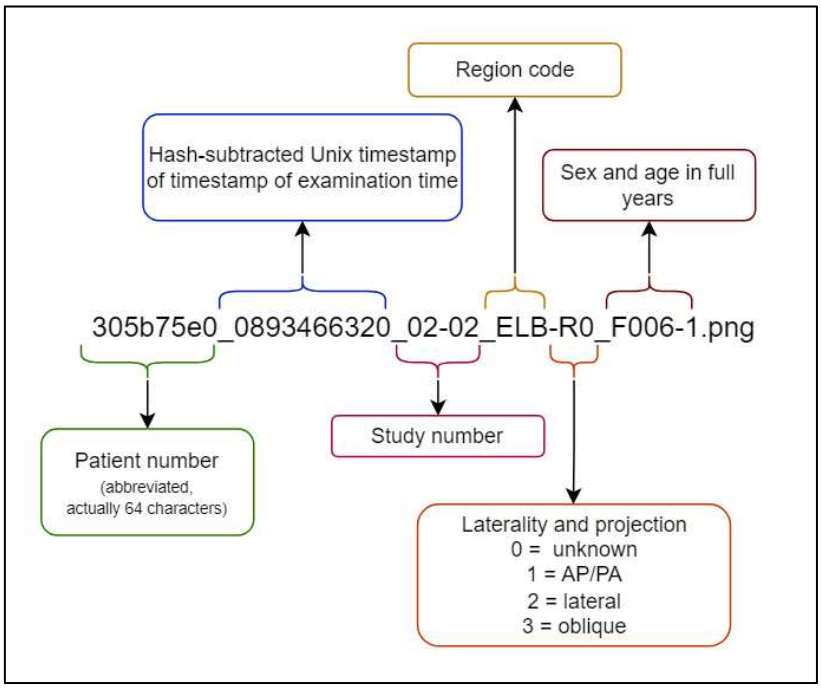

Figure 4 Meaning of the individual file name parts.

4.4. Image preprocessing

We chose to convert the DICOM images to the Portable Network Graphics (PNG) format, since PNG images provide lossless compression for pixel data and can store information in grayscale formats with more than 8 bits. To achieve that, we adjusted the grayscale values of all DICOM images to 16-bit, expanding the grayscale histogram from the original range of 0 to 4,095 (12-bit) to 0 to 65,535 (16-bit) through the “pydicom” and “cv2” packages in Python. Next, intensity rescaling was performed by cropping the lower and upper 0.05^{th} percentiles of the image histograms. Additionally, local contrast enhancement using “exposure_adapthist” with default settings was applied. After this, the images were saved as 8- bit PNG files to the hard drive for further analyses.

4.5. DICOM layers and overlays

Any additional layers or overlays in the original DICOM were irreversibly discarded during the image conversion and de-identification process.

4.6. Reports serve as ground truth for labeling

All available radiology reports were sorted into either of the following categories “fracture” or “no fracture”, forming the ground truth for labeling fractures in the X-ray images during the annotation phase.

4.7. EfficientNet’s and YOLOv8’s CNN models and their training

The primary advantage of CNNs over their predecessors is their ability to automatically identify relevant features without human supervision. The architecture of CNNs is modeled after the neurons in human and animal brains, much like a traditional neural network (69). A CNN’s architecture is made up of an input layer, multiple hidden layers and finally an output layer. One can categorize these layers into three sorts: convolutional (CONV), pooling (POOL) and fully connected (FC) layers (70).

In our study we selected 2 different and common convolutional neural networks namely YOLOv8 and EfficientNet, training them to detect elbow fractures in children.

The models were trained on computers that belong to the Medical University of Graz, equipped with the Linux operating system and two 24 GB of memory graphics card. Before training both CNNs on pediatric elbow fractures, a pretraining of both using the ImageNet dataset was performed. We did this step so that only the final layers of the network, which are crucial for making a prediction about an image's contents, are trained to become able to do so. This step is known as "transfer learning" and provides us with the advantage of cutting training time down significantly. Meanwhile, the fundamental lower layers of the CNNs remain unchanged (71).

4.7.1. EfficientNet

EfficientNet, introduced in 2019, is counted among the latest state-of the-art convolutional neural networks. It shines with achieving higher accuracy and quicker data processing while needing a much lower number of parameters compared to a lot of its predecessors (e.g. AmoebaNet, ResNet, NASNet, SENet), on publicly available comparative datasets. It is able to do so by a process called "compound scaling", which means that it scales the depth, width, and resolution of the CNN in a balanced fashion. The depth of a CNN refers to the number of layers it has. Whereas the width is defined by the number of feature maps, which are essentially bounding boxes highlighting a specific feature of the input data. Hence a wider width results in a bigger number of feature maps that can be depict more of an image's features. And finally, the resolution scaling means that the CNN can manage a wider range of pixel formats which entails it to capture finer details. So, in balancing the scales of depth, width and resolution a higher accuracy is achievable. With every more advanced variant of the EfficientNet CNN (B0 to B7, with B0 being the original variant and B7 being the most advanced), higher accuracies were achieved (72).

4.7.2. YOLOv8

In 2015 the original version of the "You-Only-Look-Once"-series, YOLOv1, was published by Redmon J. et al. (73). The "Once" aspect of the YOLO models refers to their ability to make class predictions, directly from the image's pixels by drawing bounding boxes around the region of interest (ROI) "all in one go" (hence "only looking once") and in real time. This enables them to cut processing time down

drastically and achieve higher accuracy while demanding significantly fewer parameters, in comparison to older generations which are two-staged object detectors (73). Over time, they have undergone multiple improvements in their structure and have incorporated various concepts, with YOLOv10 being the newest and most advanced (state-of-the-art) addition to the YOLO series (74).

The architecture of YOLOv8 can be broken down into several key parts (its basic architecture can be seen in Figure 5):

1. **Input layer (preprocessing and augmentation):** Before YOLO can begin to work on an object detection project, input images are usually pre-processed. This helps make the images compatible with the network and lowers the computational burden (75).
2. **Backbone network (feature extraction):** The backbone network serves as the core of YOLOv8, tasked with extracting features from the input image. YOLOv8 utilizes CSPDarknet53, a modified CNN version of Darknet, as its backbone. The CSPDarknet53 architecture features a unique Cross-Stage Partial (CSP) connection, which significantly boosts the exchange of information across various layers of the network. This connection also ensures more efficient gradient flow throughout the training process (76). Therefore, in our study, we used the CSPDarknet53 model along with the suggested improvements (referred to as the ”bag of specials”).
3. **Neck (feature aggregation):** YOLOv8 incorporates a Path Aggregation Network (PANet) as its neck architecture. This structure enhances the transmission of information between different spatial levels, allowing the model to efficiently capture features at multiple scales (76).
4. **Head (detector):** The head architecture includes several detection heads, with each one tasked with predicting bounding boxes, class probabilities, and objectness scores across various scales (76). The true innovation in YOLOv8 is found in its detection head. It features an enhanced version of the YOLO head, integrating dynamic anchor assignment and a new IoU (Intersection over Union) loss function. These advancements lead to more precise bounding box predictions and improved management of overlapping objects (76).

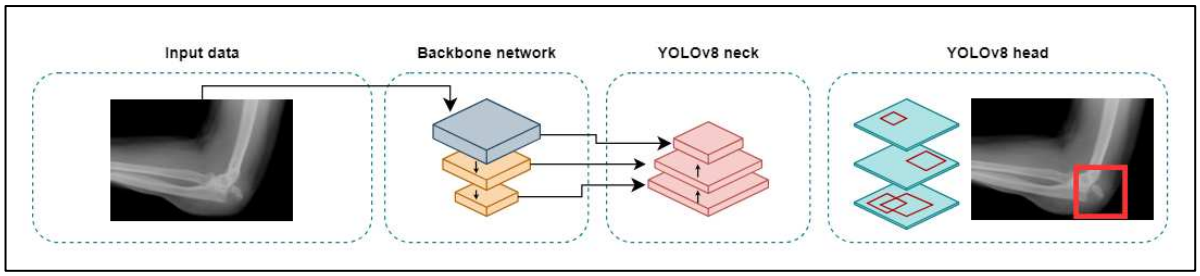

Figure 5 Basic architecture of YOLOv8.

YOLOv8 comes in 5 variants. The difference between these CNN variants lies in the number of their parameters that is the number of their learnable elements. The variants of the YOLOv8 model increase in the number of parameters as follows: nano (n), small (s), medium (m), large (l), and extra-large (x) (77). The more complex (higher number of parameters) the CNN, the higher the accuracy and the longer the computing time. Hence, the simpler the CNN, the lower the accuracy and the shorter the computing time (78). In our study, we tested all 5 variants on our test sets.

4.8. Test metrics

To be able to compare both models between themselves and with different models (or variants of the same sort), we measured the standardly widely adopted performance metrics for AI (79). Among them are the sensitivity (= true positive rate or TPR), specificity (= true negative rate or TNR), F1-score (= the harmonic mean between precision and recall or F1). These metrics are based on the so called "confusion matrix", which compares positive and negative predictions to the actual outcomes. The matrix is structured as a 2 × 2 table which comprises the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

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

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

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

$$Negative\ predictive\ value\ (NPV) = \frac{TP}{(TP + FP)}$$

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

$$F1\text{–}score = \frac{2 * (precision * recall)}{(precision + recall)} = \frac{TP}{TP + \frac{1}{2} + (FP + FN)}$$

4.9. Statistical analysis

Statistical analysis was done with IBM SPSS Statistics Version 29.0.0, mainly employing descriptive statistics. Precision-Recall Area under the Curve (PR-AUC) analysis, which shows sensitivity in relation to precision, were performed for the binary classification of fractures (presence or absence) for both CNNs.

5. Results

5.1. EfficientNet's performance

Out of all EfficientNet variants, B3 achieved the highest recall, F1-score and accuracy. In terms of precision, specificity, and NPV, B6 recorded the highest results. Table 3 includes graphs that plot the minute differences between the different variants in specific metrics (precision, recall, F1-score and accuracy). Table 3 shows a detailed overview of the results of various metrics. In Figure 6 one can

compare the sub-variants' performances regarding specific metrics with one another. Figure 7 shows PR-curves for all the utilized sub-variants of EfficientNet. A summary of all PR-curves of the utilized sub-variants of EfficientNet can be seen in Figure 8. B0 achieved a PR-AUC of 0.94, B2 and B4 have a PR-AUC of 0.95, B1, B3 and B7 have a PR-AUC of 0.96, and finally B5 and B6 have a PR-AUC of 0.97.

	Precision	Recall	F1-score	NPV	Accuracy	Specificity
EfficientNet-B0	0.8997	0.8268	0.8617	0.8997	0.81	0.7676
EfficientNet-B1	0.9025	0.9050	0.9038	0.9025	0.862	0.7535
EfficientNet-B2	0.8917	0.8743	0.8829	0.8917	0.834	0.7324
EfficientNet-B3	0.9013	0.9441	0.9222	0.9013	0.886	0.7394
EfficientNet-B4	0.8989	0.9190	0.9088	0.8989	0.868	0.7394
EfficientNet-B5	0.9066	0.9218	0.9141	0.9066	0.876	0.7606
EfficientNet-B6	0.9162	0.9162	0.9162	0.9162	0.88	0.7887
EfficientNet-B7	0.9116	0.9218	0.9167	0.9116	0.88	0.7746

Table 3 Results of the various calculated metrics for the EfficientNet sub-variants.

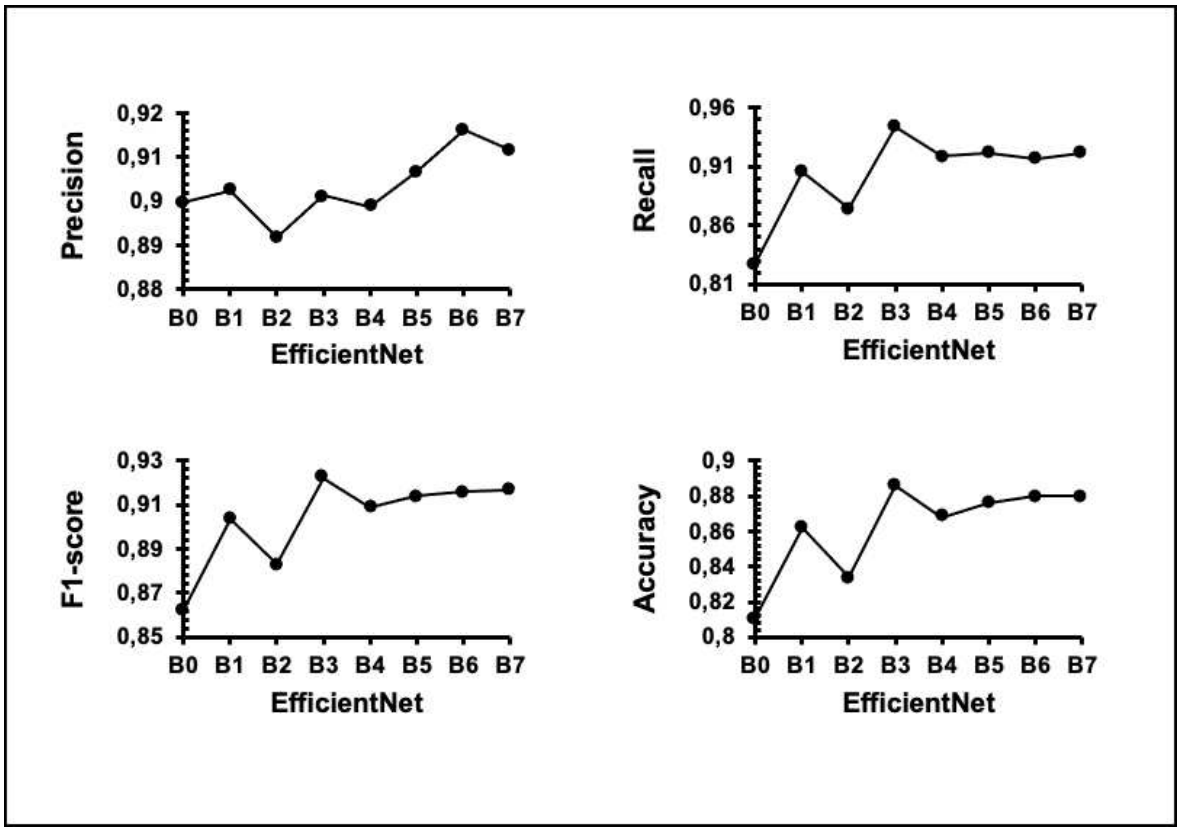

Figure 6 Comparison of different metrics between the utilized EfficientNet sub-variants.

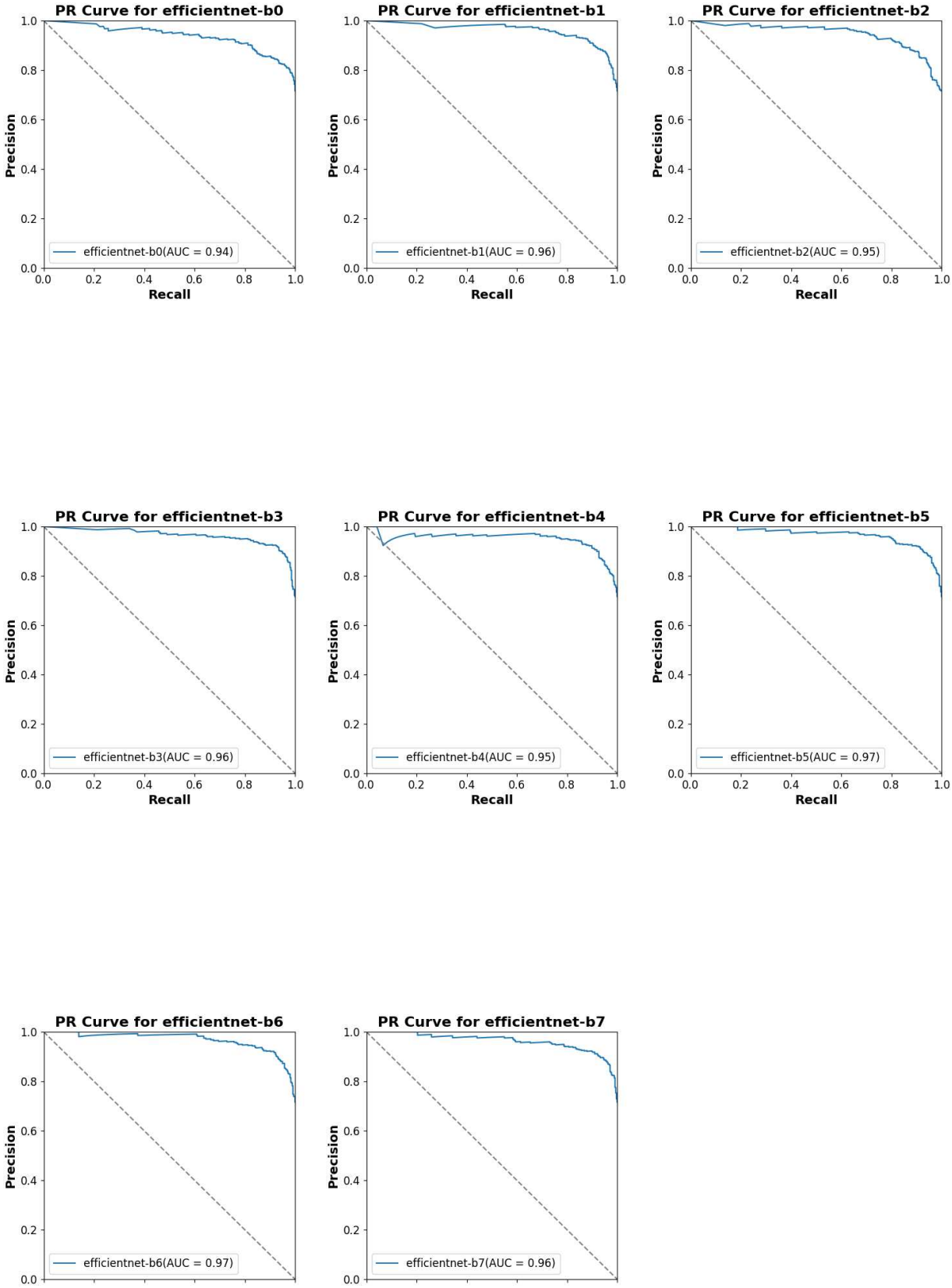

Figure 7 PR-curves for the utilized EfficientNet sub-variants.

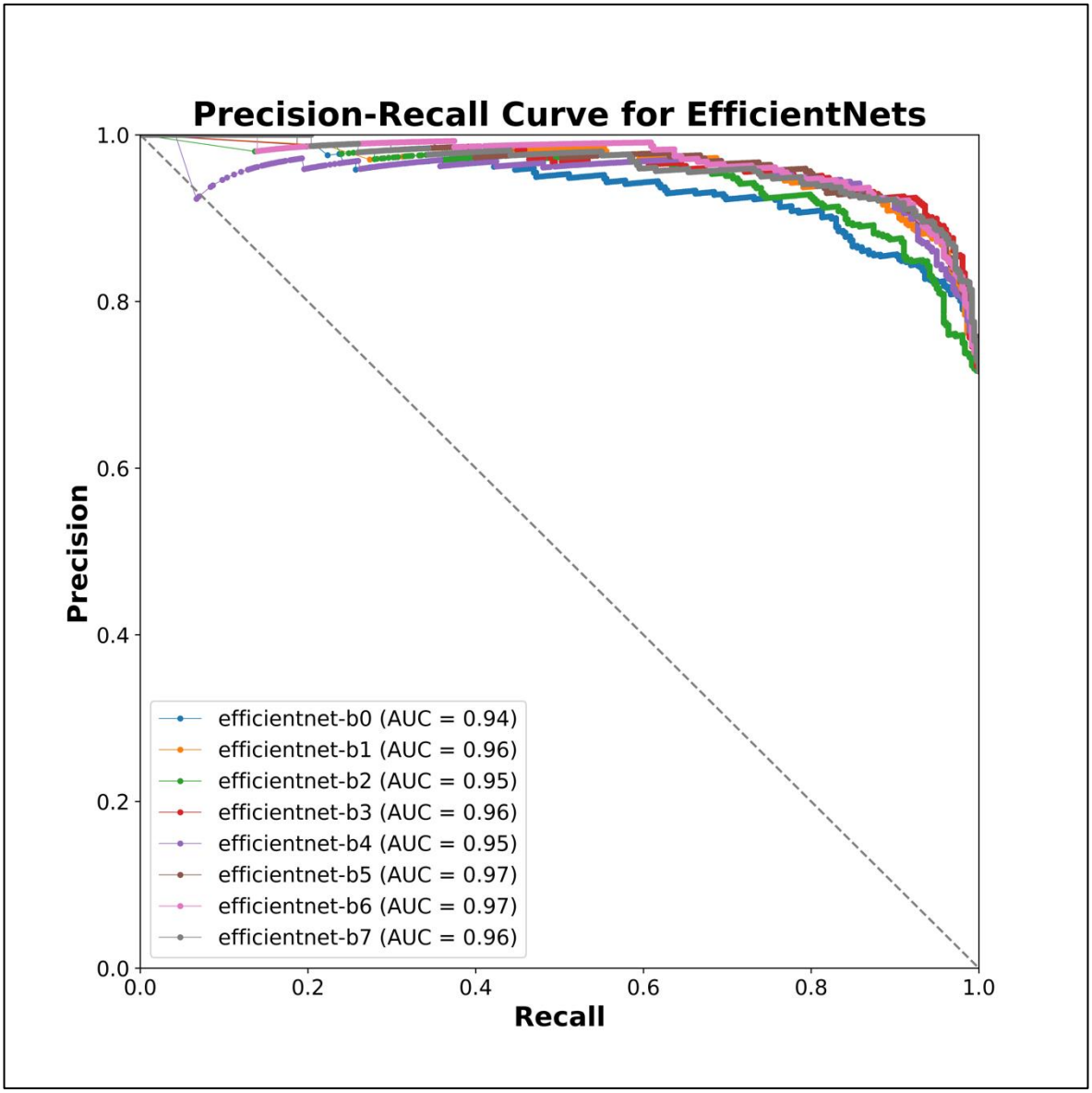

Figure 8 PR-curves of all utilized EfficientNet sub-variants summarized in one plot.

5.2. YOLOv8's performance

Between the different YOLOv8 variants, the extra-large (x) variant attained the highest result in precision, F1-score and AP50-95%. YOLOv8-n and YOLOv8-s tied for the highest score in NPV. Recall was the highest in YOLOv8-l. Table 4 shows a detailed overview of the results of various metrics. In Figure 9 one can compare the sub-variants' performances regarding specific metrics with one another. Figure 10 to Figure 14 show PR-curves for all the utilized sub-variants of YOLOv8. YOLOv8-m, YOLOv8-l, YOLOv8-x have each a PR-AUC of 0.954 whereas YOLOv8-n, and YOLOv8-s each have a PR-AUC of 0.951.

	Precision	Recall	F1-score	NPV	AP50%	AP50-95%
YOLOv8-n	0.9369	0.8811	0.9082	0.9746	0.9510	0.6468
YOLOv8-s	0.9300	0.8950	0.9122	0.9746	0.9515	0.6412
YOLOv8-m	0.9152	0.9003	0.9077	0.9693	0.9544	0.6519
YOLOv8-l	0.8985	0.9111	0.9047	0.9719	0.9536	0.6490
YOLOv8-x	0.9440	0.8868	0.9145	0.9639	0.9542	0.6612

Table 4 Results of the various calculated metrics for the YOLOv8 sub-variants.

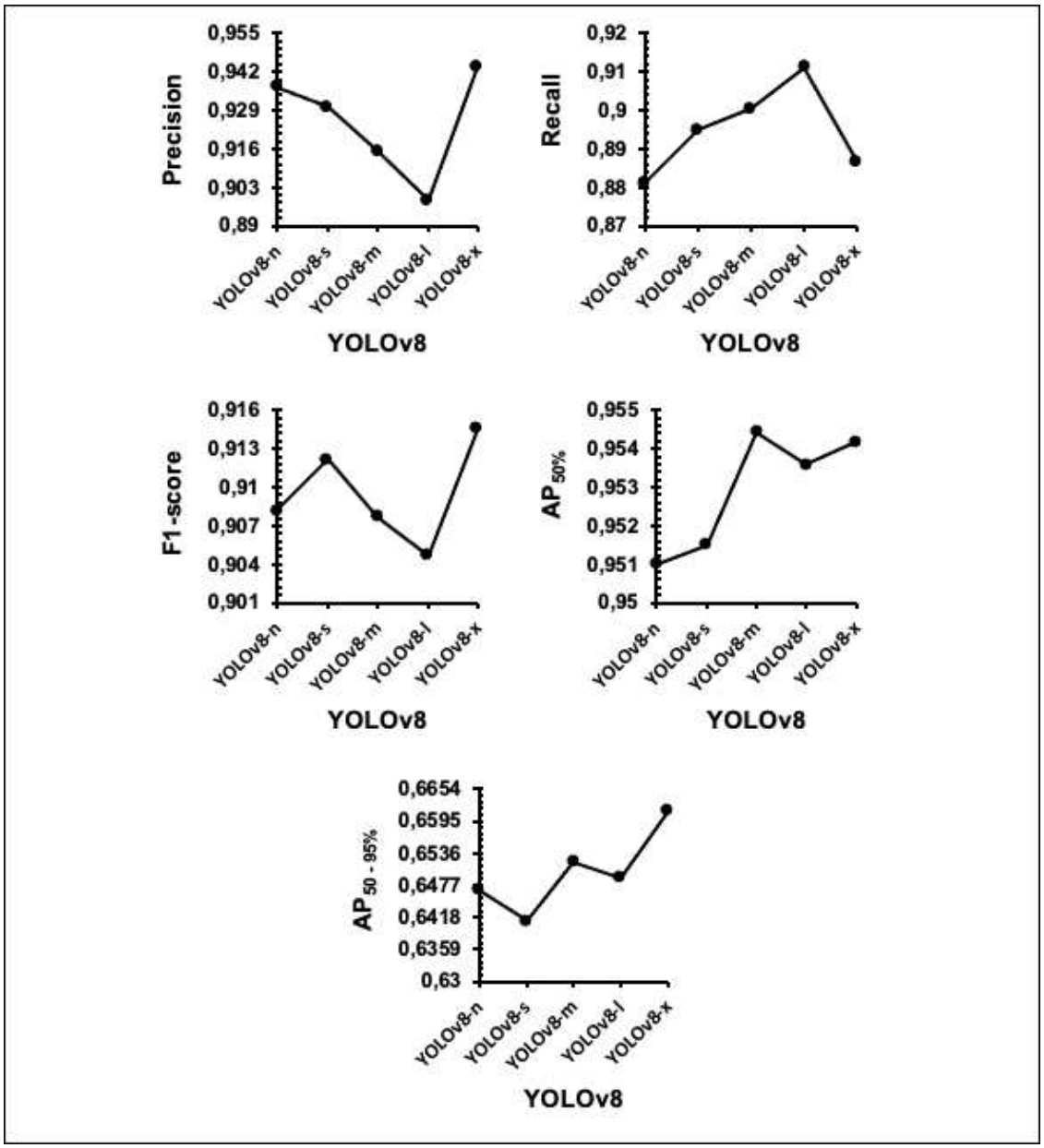

Figure 9 Comparison of different metrics between the utilized YOLOv8 sub-variants.

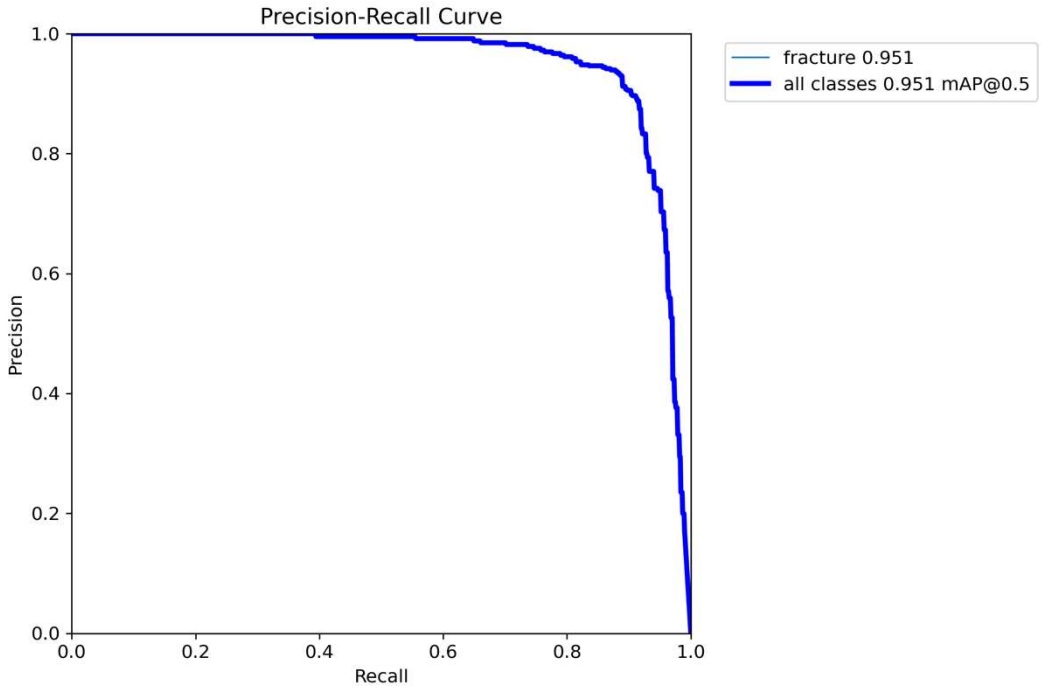

Figure 10 PR-curve of YOLOv8-n.

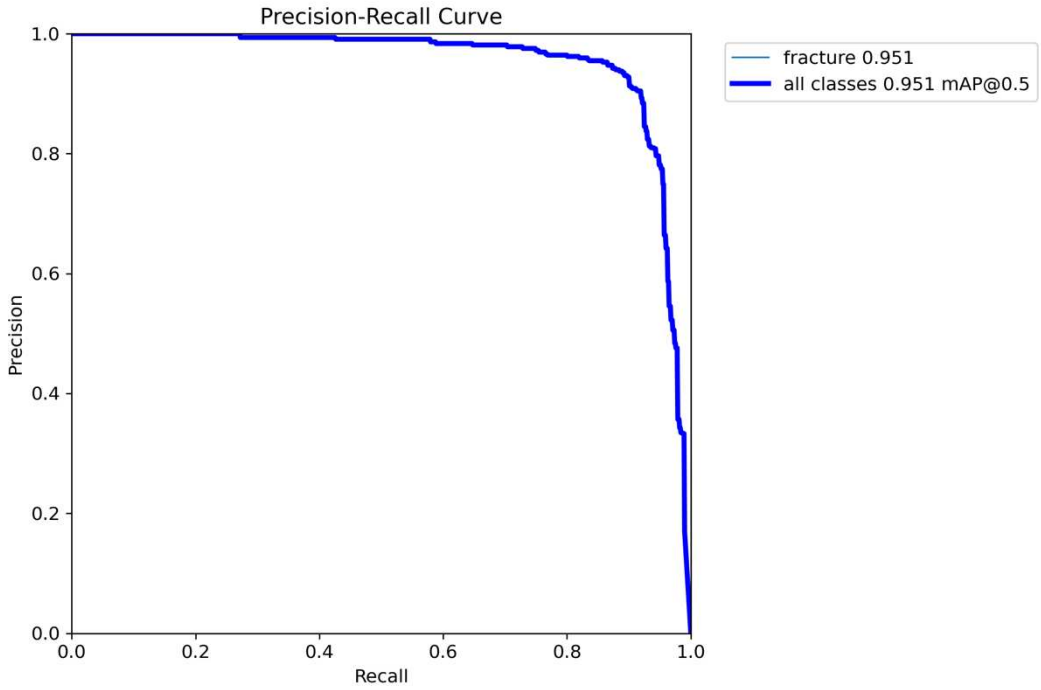

Figure 11 PR-curve of YOLOv8-s.

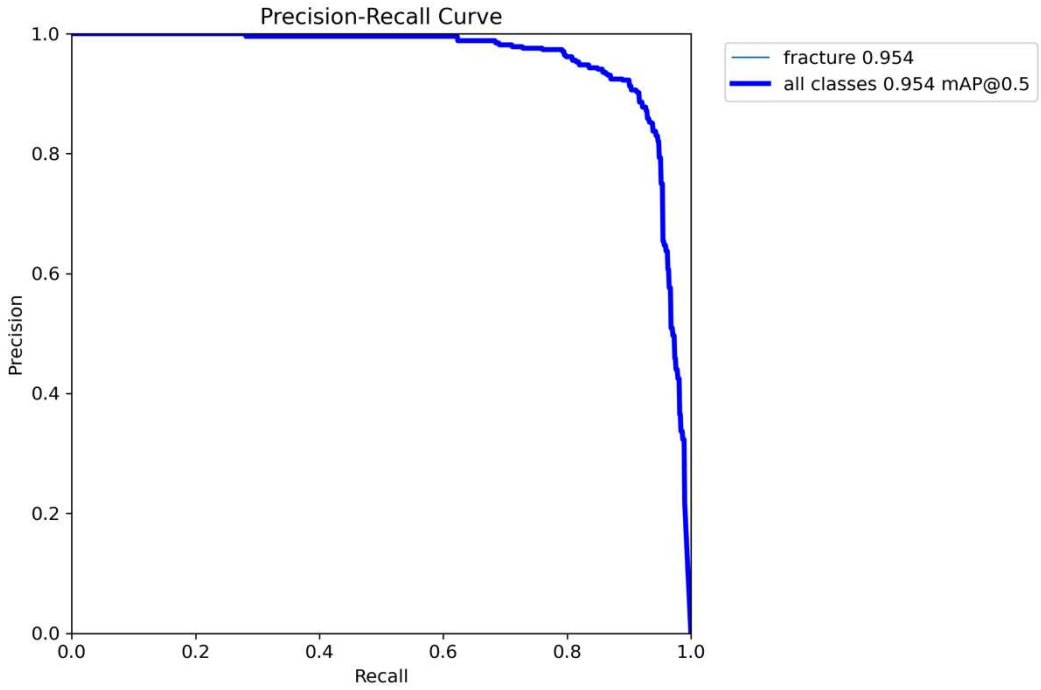

Figure 12 PR-curve of YOLOv8-m.

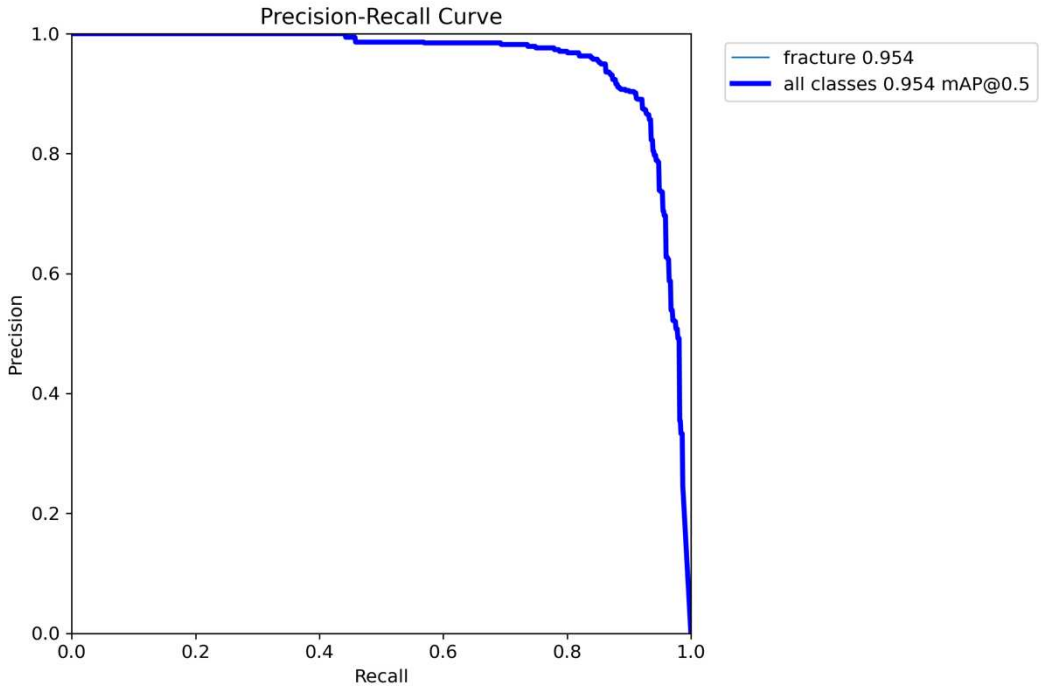

Figure 13 PR-curve of YOLOv8-l.

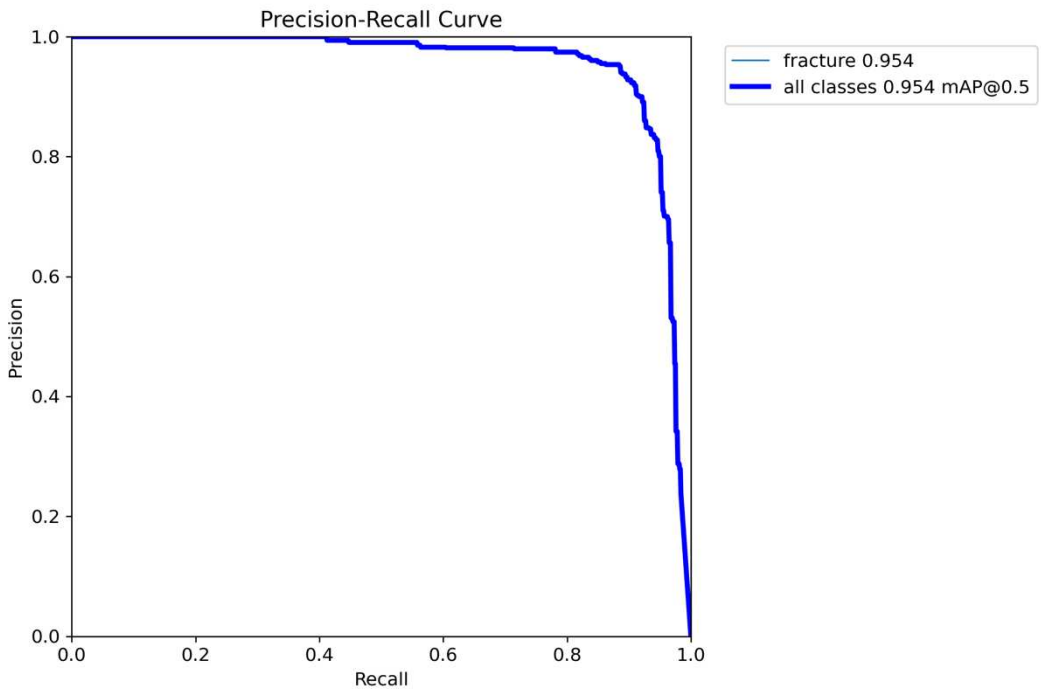

Figure 14 PR-curve of YOLOv8-x.

6. Discussion

We trained and evaluated two different computer vision model families (EfficientNet and YOLOv8) for their capability to detect and localize elbow fractures in children. We showed that for example B3 from EfficientNet achieved a recall of 94,41%, which means that out of all the fractures in the test set 94,41% could be identified as such. Considering studies like the following one reported an accuracy rate of 53% for human emergency room physicians to identify pediatric elbow fractures correctly (80), our findings suggest that our developed algorithms might be better than a substantial number of human raters.

Although in this study we achieved a recall rate of 94.41% for a certain sub-variant, meaning approximately 5.59% of fractures went undetected, this proportion of missed fractures is relatively high in a medical context. Therefore, at this point, we cannot fully rely on AI-based analysis alone for fracture detection. To put this into perspective: 5.59% of 1,000,000 fractures would amount to 55,900 undetected cases, which is an extremely high number of patients who could suffer severe complications, potentially leading to irreversible disability. Thus, while a recall rate

of 94.41% may sound promising at first, it is clear that autonomous use of AI in the diagnosis of pediatric elbow fractures is currently not feasible.

Comparing the performance of our models to those of the study described in chapter 1.3 by Dupuis M. and colleagues (17) is of limited value, because neither our models were tested on the test set used in that study nor vice versa. The same applies to the second study described in chapter 1.3 by Kekatpure A. and colleagues (18), because again neither our models were tested on the test set used in that study nor vice versa. Additionally, Kekatpure A.'s study trained their model solely on distal humerus fractures, whereas our models where trained on elbow fractures, containing not only distal humerus fractures but also proximal radius and proximal ulna fractures.

In a medical context, particularly when attempting to detect diseases, there is a tendency to prioritize AI algorithms with a relatively high recall (sensitivity), even at the cost of lower precision. The reason behind this is because false positive results are generally less severe than false negatives. In the case of a false positive, where a patient is incorrectly told they have a fracture, further examinations usually help to rule out the misdiagnosis. However, in clinical practice, a false negative—where a patient is incorrectly informed, they do not have a fracture—is far more serious. In such a scenario, the patient may go home with a false sense of security, ignoring pain or minor discomfort, which can eventually lead to more severe complications that are harder to manage. F1-scores are a common metric to compare AI models with each other because it puts both the precision and recall metric into perspective. However as just described a false negative is a lot less tolerable than a false positive. In that regard EfficientNet-B7 might be a better choice compared to EfficientNet-B3, despite EfficientNet-B3 having a higher F1-score compared to EfficientNet-B7. The reason behind that is that EfficientNet-B7 has a higher NPV and therefore has a lower false negative rate compared to EfficientNet-B3.

To further improve computer vision model performance in the future key strategies include using larger, more diverse, and balanced datasets (81). In the future, we could enhance our developed algorithms by adopting these strategies. For instance, gathering additional data from different pediatric radiology departments with an even

more diverse pediatric patient population, would allow us to create a dataset that is not only larger but also better balanced, hopefully resulting into an overall improvement of the models' performances.

6.1. Further outlook beyond this thesis

By research like this one, we contributed to improving the diagnosis of pediatric elbow fractures in radiological imaging using AI, bringing the scientific community one step closer to fulfilling its long-term goal of making AI surpass human expert accuracy. While human diagnosticians possess valuable experience and intuition, they are inherently limited by factors such as fatigue, illness, and the need for vacation time. In contrast, AI algorithms can operate continuously without suffering from these limitations, providing consistent and reliable diagnostic capabilities. This aspect is particularly crucial in high-demand healthcare settings, where timely and accurate diagnoses can significantly impact patient outcomes. Furthermore, as discussed in the chapter "Why we should care about implementing AI in diagnostic imaging," AI can help reduce variability in diagnoses and enhance the overall efficiency of the diagnostic process. By creating more dependable and precise algorithms, we helped to complement human expertise and improve the quality of care for pediatric patients with fractures. Another significant long-term subsequence of this study is to make the developed algorithms publicly accessible through a website. To be more specific, that means that our department planned to create a user-friendly platform where individuals can upload their radiographs (as DCM, PNG, JPG, BMP, TIF images) to determine whether the algorithm can accurately identify a fracture. The final version of this website is yet to be released officially. This initiative aims to provide an additional resource for patients and healthcare providers seeking a second opinion, thus enhancing accessibility to diagnostic tools.

With our project we also seek to address socioeconomic inequities in the availability of medical diagnostics. By offering a web-based tool that allows users to analyse their radiographs for potential fractures, we aim to help bridge the gap in access to healthcare resources, particularly for underserved populations. This could empower families and individuals who may not have easy access to medical specialists, thereby promoting equity in pediatric healthcare. To maintain user privacy and

comply with ethical standards, we will implement automated anonymization of uploaded radiographs. This ensures that the personal information of users is protected while allowing the algorithms to analyse images effectively. By prioritizing data security, we aim to foster trust in the platform and encourage more users to utilize the tool for their diagnostic needs.

6.2. Limitations

Limitations of the work are that the study was conducted using retrospective data from a single-center institution, which had a limited number of radiographic studies available. This inevitably introduces a selection bias and leaves the generalizability of the AI models uncertain. To minimize this bias, conducting a large-scale, multi-center study would be crucial, as it would provide a more comprehensive training dataset and significantly improve the model's performance. Further studies can build upon the generated results to overcome this limitation.

Due to the lack of follow-up radiographs or images from other modalities, such as MRI or CT, occult fractures could not be definitively ruled out. This means that some of the false positives could actually be true positives, and some false negatives could be true negatives, and vice versa. Had we excluded all ambiguous cases from the outset, we could have minimized the number of incorrectly classified ground truths. Still, there is some noise within the dataset.

As mentioned in the introduction, black boxes remain a significant challenge in explaining the decision-making process of AI. Selvaraju and colleagues (82) introduced a technique called "Grad-CAM", which highlights the key pixels that most influenced the AI's decision-making process in the form of a color-coded attention heat map. The major advantage of this method is that Grad-CAM works with any type of CNN without requiring modifications to the CNN itself. Unfortunately, we were unable to integrate this technique into our study, which means we missed the deeper insight into the black box that we could have otherwise achieved.

6.3. Conclusions

In conclusion this thesis supports the idea that machine learning has potential to support medical experts while managing pediatric elbow fractures. Both EfficientNet

and YOLOv8 models demonstrated strong potential for assisting in the detection of pediatric elbow fractures, though their performance varied depending on specific metrics. Our findings underscore the potential for AI models to enhance radiographic diagnosis in pediatric settings, though further refinement and clinical validation are necessary to ensure their effectiveness in real-world applications.

VIII. References

1. Xing L, Giger ML, Min JK, editors. Artificial intelligence in medicine: technical basis and clinical applications. London, United Kingdom ; San Diego, CA: Academic Press, an imprint of Elsevier; 2020. 544 p.

2. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol. 2019 Mar 4;28(2):73–81.

3. Mrs. C. Radha, Mr. R. Midunkumar, Mr. S. Muralibabu, Mr. V. Partheeban, Mr. V. Partheeban. Role of Artificial Intelligence in Big Data Analytics. Int J Adv Res Sci Commun Technol. 2024 Apr 11;586–91.

4. Rasouli Panah H. Early Detecting of Infectious Disease Outbreaks: AI Potentials for Public Health Systems. Rangahau Aranga AUT Grad Rev [Internet]. 2023 Oct 2 [cited 2024 Jul 20];2(3). Available from: https://ojs.aut.ac.nz/rangahau-aranga/article/view/180

5. - TRW, - MSR. Revolutionizing Radiology: Exploring Applications and Advancements in AI for Imaging Diagnostics. Int J Multidiscip Res. 2023 Dec 9;5(6):10291.

6. Manning D, Ethell S, Donovan T, Crawford T. How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography. 2006 May;12(2):134–42.

7. Zhang J, Wang Y, Yu B, Shi X, Zhang Y. Application of Computer-Aided Diagnosis to the Sonographic Evaluation of Cervical Lymph Nodes. Ultrason Imaging. 2016 Mar;38(2):159–71.

8. Dheeba J, Albert Singh N. Computer Aided Intelligent Breast Cancer Detection: Second Opinion for Radiologists—A Prospective Study. In: Azar AT, Vaidyanathan S, editors. Computational Intelligence Applications in Modeling and Control [Internet]. Cham: Springer International Publishing; 2015 [cited 2024 Oct 24]. p. 397–430. (Studies in Computational Intelligence; vol. 575). Available from: https://link.springer.com/10.1007/978-3-319-11017-2_16

9. Harada T, Shimizu T, Kaji Y, Suyama Y, Matsumoto T, Kosaka C, et al. A Perspective from a Case Conference on Comparing the Diagnostic Process: Human Diagnostic Thinking vs. Artificial Intelligence (AI) Decision Support Tools. Int J Environ Res Public Health. 2020 Aug 22;17(17):6110.

10. Bhargavan M, Sunshine JH, Schepps B. Too Few Radiologists? Am J Roentgenol. 2002 May;178(5):1075–82.

11. Bruls RJM, Kwee RM. Workload for radiologists during on-call hours: dramatic increase in the past 15 years. Insights Imaging. 2020 Dec;11(1):121.

12. Carabantes M. Black-box artificial intelligence: an epistemological and critical analysis. AI Soc. 2020 Jun;35(2):309–17.

13. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020 Jun;58:82–115.

14. Janisch M, Apfaltrer G, Hržić F, Castellani C, Mittl B, Singer G, et al. Pediatric radius torus fractures in x-rays—how computer vision could render lateral projections obsolete. Front Pediatr. 2022 Dec 14;10:1005099.

15. Chacon D, Kissoon N, Brown T, Galpin R. Use of comparison radiographs in the diagnosis of traumatic injuries of the elbow. Ann Emerg Med. 1992 Aug;21(8):895–9.

16. Currie G, Rohren E. Social Asymmetry, Artificial Intelligence and the Medical Imaging Landscape. Semin Nucl Med. 2022 Jul;52(4):498–503.

17. Dupuis M, Delbos L, Rouquette A, Adamsbaum C, Veil R. External validation of an artificial intelligence solution for the detection of elbow fractures and joint effusions in children. Diagn Interv Imaging. 2024 Mar;105(3):104–9.

18. Kekatpure A, Kekatpure A, Deshpande S, Srivastava S. Development of a diagnostic support system for distal humerus fracture using artificial intelligence. Int Orthop. 2024 May;48(5):1303–11.

19. Otto S, Wiersbitzky H, Mutze S, Hosten N. Frakturen bei Kindern und Jugendlichen. Radiol Up2date. 2003 Sep;3(3):257–83.

20. Riccabona M, Lindbichler F. Traumaradiologie beim Kind. Radiol. 2002 Mar;42(3):195–209.

21. Sharma R, Jones J. Incomplete fracture. In: Radiopaedia.org [Internet]. Radiopaedia.org; 2013 [cited 2024 Jul 21]. Available from: http://radiopaedia.org/articles/22210

22. Niknejad M, Weerakkody Y. Torus fracture. In: Radiopaedia.org [Internet]. Radiopaedia.org; 2010 [cited 2024 Jul 23]. Available from: http://radiopaedia.org/articles/10800

23. Saber M, Radswiki T. Greenstick fracture. In: Radiopaedia.org [Internet]. Radiopaedia.org; 2010 [cited 2024 Jul 23]. Available from: http://radiopaedia.org/articles/12606

24. McCoy JS, Nelson R. Avulsion Fractures. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Jul 23]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK559168/

25. Merrow AC, Reiter MP, Zbojniewicz AM, Laor T. Avulsion fractures of the pediatric knee. Pediatr Radiol. 2014 Nov;44(11):1436–45.

26. Merckaert S, Chaibi E, Meriem S, Kwiatkowski B, Divjak N, Zambelli PY. Epidemiology of Pediatric Upper Extremity Fractures in a Tertiary Care Center in Switzerland. Pediatr Emerg Care. 2021 Dec;37(12):e825–35.

27. Bryson DJ, Shivji FS, Price KR, Lawniczak D, Chell J, Hunter JB. The lost art of conservative management of paediatric fractures. Bone Jt 360. 2016 Feb;5(1):2–8.

28. Pusic M, Pecaric M, Boutis K. How Much Practice Is Enough? Using Learning Curves to Assess the Deliberate Practice of Radiograph Interpretation: Acad Med. 2011 Jun;86(6):731–6.

29. Dawes TJW, Vowler SL, Allen CMC, Dixon AK. Training improves medical student performance in image interpretation. Br J Radiol. 2004 Sep;77(921):775–6.

30. Martino F, Barbuti D, Martino G, Cirillo M. Missed Fractures in Children. In: Romano L, Pinto A, editors. Errors in Radiology [Internet]. Milano: Springer Milan; 2012 [cited 2024 Jul 31]. p. 51–63. Available from: http://link.springer.com/10.1007/978-88-470-2339-0_6

31. Kraus R, Dresing K. Rational Usage of Fracture Imaging in Children and Adolescents. Diagnostics. 2023 Feb 1;13(3):538.

32. Schmid GL, Lippmann S, Unverzagt S, Hofmann C, Deutsch T, Frese T. The Investigation of Suspected Fracture— a Comparison of Ultrasound With Conventional Imaging. Dtsch Ärztebl Int [Internet]. 2017 Nov 10 [cited 2024 May 21]; Available from: https://www.aerzteblatt.de/10.3238/arztebl.2017.0757

33. De Zwart AD, Beeres FJP, Rhemrev SJ, Bartlema K, Schipper IB. Comparison of MRI, CT and bone scintigraphy for suspected scaphoid fractures. Eur J Trauma Emerg Surg. 2016 Dec;42(6):725–31.

34. Moritz J, Hoffmann B, Sehr D, Eggerking J, Keil K, Caliebe A, et al. Pädiatrische Frakturdiagnostik – Ultra-low-dose-CT mit der effektiven Dosis von Röntgenaufnahmen. RöFo - Fortschritte Auf Dem Geb Röntgenstrahlen Bildgeb Verfahr. 2012 Aug 7;184(11):1026–33.

35. Kolk A, Stimmer H, Klopfer M, Wolff KD, Hohlweg-Majert B, Ploder O, et al. High Resolution Magnetic Resonance Imaging With an Orbital Coil as an Alternative to Computed Tomography Scan as the Primary Imaging Modality of Pediatric Orbital Fractures. J Oral Maxillofac Surg. 2009 Feb;67(2):348–56.

36. Orman G, Wagner MW, Seeburg D, Zamora CA, Oshmyansky A, Tekes A, et al. Pediatric skull fracture diagnosis: should 3D CT reconstructions be added as routine imaging? J Neurosurg Pediatr. 2015 Oct;16(4):426–31.

37. Franklin DB, Hardaway AT, Sheffer BW, Spence DD, Kelly DM, Muhlbauer MS, et al. The Role of Computed Tomography and Magnetic Resonance Imaging in the Diagnosis of Pediatric Thoracolumbar Compression Fractures. J Pediatr Orthop. 2019 Aug;39(7):e520–3.

38. Dodd A, Oddone Paolucci E, Korley R. The effect of three-dimensional computed tomography reconstructions on preoperative planning of tibial plateau fractures: a case–control series. BMC Musculoskelet Disord. 2015 Dec;16(1):144.

39. Frush DP, Donnelly LF, Rosen NS. Computed Tomography and Radiation Risks: What Pediatric Health Care Providers Should Know. Pediatrics. 2003 Oct 1;112(4):951–7.

40. Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. BMJ. 2013 May 21;346(may21 1):f2360–f2360.

41. Wilson MP, Nobbee D, Murad MH, Dhillon S, McInnes MDF, Katlariwala P, et al. Diagnostic Accuracy of Limited MRI Protocols for Detecting Radiographically Occult Hip Fractures: A Systematic Review and Meta-Analysis. Am J Roentgenol. 2020 Sep;215(3):559–67.

42. Johnson B, Alizai H, Dempsey M. Fast field echo resembling a CT using restricted echo-spacing (FRACTURE): a novel MRI technique with superior bone contrast. Skeletal Radiol. 2021 Aug;50(8):1705–13.

43. Schwaiger BJ, Schneider C, Kronthaler S, Gassert FT, Böhm C, Pfeiffer D, et al. CT-like images based on T1 spoiled gradient-echo and ultra-short echo time MRI sequences for the assessment of vertebral fractures and degenerative bone changes of the spine. Eur Radiol. 2021 Jul;31(7):4680–9.

44. Booz C, Nöske J, Albrecht MH, Lenga L, Martin SS, Bucher AM, et al. Diagnostic accuracy of color-coded virtual noncalcium dual-energy CT for the assessment of bone marrow edema in sacral insufficiency fracture in comparison to MRI. Eur J Radiol. 2020 Aug;129:109046.

45. Iacob R, Stoicescu ER, Cerbu S, Iacob D, Amaricai E, Catan L, et al. Could Ultrasound Be Used as a Triage Tool in Diagnosing Fractures in Children? A Literature Review. Healthcare. 2022 Apr 29;10(5):823.

46. Moritz JD. Sonographic Fracture Diagnosis in Children and Adolescents. RöFo - Fortschritte Auf Dem Geb Röntgenstrahlen Bildgeb Verfahr. 2023 Sep;195(09):790–6.

47. Nicholson JA, Tsang STJ, MacGillivray TJ, Perks F, Simpson AHRW. What is the role of ultrasound in fracture management?: Diagnosis and therapeutic potential for fractures, delayed unions, and fracture-related infection. Bone Jt Res. 2019 Jul;8(7):304–12.

48. Bawale R, Segmeister M, Sinha S, Shariff S, Singh B. Experience of an isolated use of low-intensity pulsed ultrasound therapy on fracture healing in established non-unions: a prospective case series. J Ultrasound. 2021 Sep;24(3):249–52.

49. Abzug JM, Herman MJ, Kozin S. Pediatric Elbow Fractures: A Clinical Guide to Management. Springer; 2018. 257 p.

50. Bisset GS, Crowe J. Diagnostic errors in interpretation of pediatric musculoskeletal radiographs at common injury sites. Pediatr Radiol. 2014 May;44(5):552–7.

51. Brodeur AE, Silberstein MJ, Graviss ER. Radiology of the pediatric elbow. Boston, Mass: G. K. Hall Medical Publishers; 1981. 234 p.

52. Kashayi-Chowdojirao S, Yadala D. Pediatric elbow – Developmental and radiological anatomy. J Orthop Assoc South Indian States. 2022;19(3):9.

53. Mattijssen-Horstink L, Langeraar JJ, Mauritz GJ, van der Stappen W, Baggelaar M, Tan ECTH. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis. Scand J Trauma Resusc Emerg Med. 2020 May 13;28(1):38.

54. Rollins JH, Long BW, Curtis T. Merrill's Atlas of Radiographic Positioning and Procedures - 3-Volume Set - E-Book: Merrill's Atlas of Radiographic Positioning and Procedures - 3-Volume Set - E-Book. Elsevier Health Sciences; 2022. 1702 p.

55. Egol KA, Koval KJ, Zuckerman JD. Handbook of Fractures. Lippincott Williams & Wilkins; 2010. 812 p.

56. Sinha A, Sapra D, Srivastava G, Anup M, Sinwar D. AI-Assisted Big Data Analytics for Smart Healthcare Systems. In: Intelligent Internet of Things for Smart Healthcare Systems [Internet]. 1st ed. Boca Raton: CRC Press; 2023 [cited 2024 Jul 18]. p. 81–100. Available from: https://www.taylorfrancis.com/books/9781003326182/chapters/10.1201/9781003326182-6

57. Reenu Maria Binoy, Parvathy Prakash, Dayana Thomas, Devika R Nath, Gigi Joseph. Deep Learning-Based Intelligent Analysis of Medical Big Data. Int J Eng Technol Manag Sci. 2022 Sep 28;6(5):727–33.

58. Kang DY, Duong HP, Park JC. Application of Deep Learning in Dentistry and Implantology. Korean Acad Oral Maxillofac Implantol. 2020 Sep 1;24(3):148–81.

59. Rao BNK, Pranitha B, Reddy BK, Varsha D, Reddy NN. Bone Fracture Detection and Classification Using Deep Learning Techniques: In: Khang A, editor. Advances in Medical Diagnosis, Treatment, and Care [Internet]. IGI Global; 2024 [cited 2024 Jul 18]. p. 92–101. Available from: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-3679-3.ch005

60. Prijs J, Liao Z, To MS, Verjans J, Jutte PC, Stirler V, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN). Eur J Trauma Emerg Surg. 2023 Apr;49(2):1057–69.

61. Hržić F, Štajduhar I, Tschauner S, Sorantin E, Lerga J. Local-Entropy Based Approach for X-Ray Image Segmentation and Fracture Detection. Entropy. 2019 Mar 28;21(4):338.

62. Shah T. About Train, Validation and Test Sets in Machine Learning [Internet]. 2017. Available from: https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

63. Elyan E, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-Garcia CF, et al. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artif Intell Surg [Internet]. 2022 Mar 22 [cited 2024 Aug 2];2. Available from: https://rgu-repository.worktribe.com/output/1631673

64. Wei CJ, Tsai WC, Tiu CM, Wu HT, Chiou HJ, Chang CY. Systematic analysis of missed extremity fractures in emergency radiology. Acta Radiol. 2006 Sep;47(7):710–7.

65. Graham RNJ, Perriss RW, Scarsbrook AF. DICOM demystified: A review of digital file formats and their use in radiological practice. Clin Radiol. 2005 Nov;60(11):1133–40.

66. Larobina M, Murino L. Medical Image File Formats. J Digit Imaging. 2014 Apr;27(2):200–6.

67. Aumasson JP, Neves S, Wilcox-O'Hearn Z, Winnerlein C. BLAKE2: Simpler, Smaller, Fast as MD5. In: Jacobson M, Locasto M, Mohassel P, Safavi-Naini R, editors. Applied Cryptography and Network Security [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013 [cited 2024 Aug 14]. p. 119–35. (Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, et al., editors. Lecture Notes in Computer Science; vol. 7954). Available from: http://link.springer.com/10.1007/978-3-642-38980-1_8

68. Dworkin MJ. SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions [Internet]. National Institute of Standards and Technology; 2015 Jul [cited 2024 Aug 14] p. NIST FIPS 202. Report No.: NIST FIPS 202. Available from: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.202.pdf

69. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021 Mar 31;8(1):53.

70. Sakib S, Ahmed N, Kabir AJ, Ahmed H. An Overview of Convolutional Neural Network: Its Architecture and Applications [Internet]. Preprints; 2019 [cited 2024 Aug 15]. Available from: https://www.preprints.org/manuscript/201811.0546/v4

71. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging. 2016 May;35(5):1285–98.

72. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019 [cited 2024 Aug 16]; Available from: https://arxiv.org/abs/1905.11946

73. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection [Internet]. arXiv; 2015 [cited 2024 Aug 18]. Available from: https://arxiv.org/abs/1506.02640

74. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. YOLOv10: Real-Time End-to-End Object Detection [Internet]. arXiv; 2024 [cited 2024 Aug 18]. Available from: http://arxiv.org/abs/2405.14458

75. Shahriar MT, Li H. A Study of Image Pre-processing for Faster Object Recognition [Internet]. arXiv; 2020 [cited 2024 Aug 24]. Available from: http://arxiv.org/abs/2011.06928

76. Torres J. YOLOv8 Architecture: A Deep Dive into its Architecture - YOLOv8 [Internet]. 2024 [cited 2024 Aug 24]. Available from: https://yolov8.org/yolov8-architecture/

77. Boesch G. viso.ai. 2023 [cited 2024 Aug 25]. A Guide to YOLOv8 in 2024. Available from: https://viso.ai/deep-learning/yolov8-guide/

78. Abdullah M. YOLO Working principle, difference between its ddifferent Variants and Versions [Internet]. Medium. 2023 [cited 2024 Aug 25]. Available from: https://medium.com/@muhabd51/yolo-working-principle-difference-between-its-ddifferent-variants-and-versions-95b8ad7b95ab

79. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep. 2024 Mar 13;14(1):6086.

80. Shrader MW, Campbell MD, Jacofsky DJ. Accuracy of Emergency Room Physicians' Interpretation of Elbow Fractures in Children. Orthopedics. 2008 Dec;31(12):1177.

81. Sundaram S, Hulkund N. GAN-based Data Augmentation for Chest X-ray Classification [Internet]. arXiv; 2021 [cited 2024 Oct 31]. Available from: https://arxiv.org/abs/2107.02970

82. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: 2017 IEEE International Conference on Computer Vision (ICCV) [Internet]. Venice: IEEE; 2017 [cited 2024 Oct 22]. p. 618–26. Available from: http://ieeexplore.ieee.org/document/8237336/