

Diplomarbeit

**Automatische Erkennung der patello - femoralen
Trochleadysplasie in digitalen Röntgenaufnahmen
mittels Maschinenlernen – eine retrospektive Studie**

eingereicht von

Andreas Pritsch

zur Erlangung des akademischen Grades

Doktor der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt an der Klinischen Abteilung für Kinderradiologie,
Universitätsklinik für Radiologie, Medizinische Universität Graz
Universitätsklinik für Orthopädie und Traumatologie,
Sektion Kinder – und Jugendorthopädie, Medizinische Universität Graz
unter der Anleitung von Univ.Ass. PD DDr. Sebastian Tschauner,
Ass. Prof. PD. Dr. Tanja Kraus

Graz, am 30.11.2023

Eidesstattliche Erklärung

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

Graz, am 30.11.2023

Andreas Pritsch eh.

Danksagungen

Zuallererst möchte ich meinen beiden Betreuenden Univ.Ass. PD Dr. Sebastian Tschauner sowie Ass.Prof. PD Dr. Tanja Kraus für die großartige Betreuung, den kollegialen Umgang während dieser Zeit und ganz besonders dafür danken, dass Ihr euch immer Zeit genommen habt, sobald Fragen meinerseits aufgekommen sind. Vielen herzlichen Dank.

Zudem möchte ich auch meinen Eltern sowie meinem Bruder danken, da sie in dieser durchaus herausfordernden Studienzeit sowie in allen anderen Lebensphasen für mich da sind und mich immer unterstützen. Ich danke euch von ganzem Herzen.

Desweiteren möchte ich meinen langjährigen Freunden aus meiner Heimat, sowie den in der Studienzeit gewonnenen Freunden in Graz für eure Aufmunterungen und eure Unterstützung danken.

Ganz besonders möchte ich meinem großartigen Kollegen und Freund Niklas danken, der gemeinsam mit mir das Studium absolvierte und in dieser Zeit einen wesentlichen Teil zu meinem Erfolg beigetragen hat.

Außerdem möchte ich mich herzlich bei meiner Partnerin bedanken, welche mir eine großartige Stütze ist. Sie brachte sehr viel Geduld, Liebe und Verständnis auf und begleitete mich ganz wundervoll auf diesem Weg.

Zusammenfassung in Deutsch

Einleitung:

Die femorale Trochleadysplasie ist einer der gewichtigsten Risikofaktoren für die Patellaluxation. Betroffen sind vor allem Jugendliche im Zusammenhang mit sportlicher Betätigung. Bei einer Patellaluxation kann mittels Röntgenbild eine Trochleadysplasie diagnostiziert und mittels Dejour Klassifikation eingeteilt werden.

Einige Convolutional Neural Network - Modelle konnten in der Vergangenheit bereits Meilensteine in Bezug auf Mustererkennung in Bilddaten beziehungsweise Unterscheidung zwischen pathologischen und nicht pathologischen Bildern setzen. Deep learning Modelle könnten daher auch bei der Erkennung einer femoropatellaren Dysplasie eine wichtige Rolle spielen.

Methode:

Retrospektiv wurden am Universitätsklinikum Graz, Abteilung für Orthopädie und Traumatologie, Sektion Kinder – und Jugendorthopädie Röntgenbilder von Patient*innen mit diagnostizierten Patellaluxationen ausgehoben. Die Bilder wurden anonymisiert und anschließend nach Dejour A-D klassifiziert. Zusätzlich wurde eine Vergleichsdatenbank aus kindertraumatologischen Röntgenbildern angefertigt, die keine Trochleadysplasie aufwies („N“). Beide Datenbanken wurden für das Training von zwei verschiedenen AI-Modellen (EfficientNet B4, YOLOv5) verwendet.

Ergebnis:

EfficientNet B4: Im Testdurchlauf konnten 98,75% der Bilder der Gruppe N (keine Trochleadysplasie) als nicht pathologisch klassifizieren, jedoch ergab die Unterscheidung der einzelnen Subtypen A-D nach Dejour weniger hohe Vorhersagewahrscheinlichkeiten.

YOLOv5 Objekterkennung: Dabei wurden in der Gruppe Control (keine Trochleadysplasie) 97% der präsentierten Knie Röntgenbilder korrekt zugeordnet, die Unterscheidung der einzelnen Subgruppen Dejour A-D führte aber zu etwas schwächeren

Vorhersageraten zwischen 5% und 60%. Bezogen auf die Gesamtheit einer femoralen Trochleopathie (Dejour A-D) konnte mit YOLOv5 mit 90% Wahrscheinlichkeit die Pathologie erkennen und der richtigen Gruppe zuordnen. Die Gruppe Negative Control erreichte einen Wert von 91% Vorhersagewahrscheinlichkeit.

Diskussion:

CNN-Algorithmen konnten in unterschiedlich präziser Art und Weise pathologische von nicht pathologischen seitlichen Knie Röntgenbildern unterscheiden. Verschiedene Parameter beeinflussen aber die Art und Weise, wie genau die Analyse der Röntgenbilder durchgeführt werden kann. Eine geringe Stichprobe führt einerseits zu einer Abnahme der Genauigkeit, sowie auch die zum Teil nicht streng seitlich eingestellten Röntgenbildaufnahmen. Jedoch konnte in dieser Arbeit gezeigt werden, dass es möglich ist, Algorithmen für die Unterscheidung der femoropatellaren Trochleadysplasie zu trainieren, um mit ihnen automatisierte Beurteilungen vornehmen zu lassen.

Abstract in English

Introduction:

Femoral trochlear dysplasia is one of the main risk factors for patellar dislocation, typically observed in pediatric and adolescent patients engaged in sports activities. In case of a dislocated patella, knee x-rays can show trochlear dysplasia by means of the Dejour classification. Recent advancements in artificial intelligence (AI) and computer vision could be a viable option to detect femoral trochlear dysplasia from X-ray data automatically.

Materials and Methods:

X-rays from patients with patellar dislocation were collected retrospectively from the Medical University Graz, Department of Orthopedics and Traumatology, Section of Pediatric Orthopedics. The collected X-rays were anonymized and categorized into Dejour classification groups A to D. A separate database containing trauma X-rays without trochlear dysplasia ("N") was set as a reference. Two distinct convolutional neural network models, EfficientNet B4 (classifier) and YOLOv5 (object detector), were trained using these two databases.

Results:

EfficientNet B4 correctly classified 98.75% of X-rays from the "N" group (lacking trochlear dysplasia) as non-pathological. However, the accuracy in classifying X-rays into groups A to D was lower as in the "N" group.

YOLOv5 achieved a 97% accuracy in classifying X-rays into the "Control" group (lacking trochlear dysplasia). However, accuracy varied considerably in the groups Dejour A to D, ranging from 5% to 60%. Overall, 90% of X-rays with trochlear dysplasia were correctly classified, while the negative control group resulted in a 91% prediction rate.

Discussion:

AI algorithms exhibited the ability to classify pathological X-rays into Dejour classification groups with varying prediction rates. A set of factors influenced the accuracy of these classifications. One main factor was the limited availability of X-rays in Dejour groups A to D, which impacted the algorithm's training and classification performance.

Another factor was the quality of the knee positioning during the X-ray examination, as suboptimal alignment can affect accuracy. Overall, given enough samples, it seems possible to train AI algorithms to automatically classify knee x-rays along Dejour classification from severity rates A-D.

Inhaltsverzeichnis

Danksagungen	i
Zusammenfassung in Deutsch	ii
Abstract in English	iv
Abkürzungen und deren Erklärung	viii
Abbildungsverzeichnis	x
Tabellenverzeichnis	xi
Einleitung - Grundlagen	1
1.1 Kniegelenk (Art. genu)	1
1.1.1 Anatomie des Kniegelenks.....	1
1.1.2 Knochen, Knorpel und Capsula synovialis	2
1.1.3 Muskulatur und Weichteilanteile	3
1.1.4 Menisci	5
1.1.5 Bandstrukturen.....	6
1.2 Trochleadysplasie	8
1.2.1 Definition	8
1.2.2 Epidemiologie der Trochleadysplasie	11
1.2.3 Diagnostik	11
1.2.4 Therapie.....	15
1.2.5 Convolutional Neural Network (CNN)	16
2 Material und Methoden	17
2.1 Datenanonymisierung	18
2.2 Röntgenbild Selektion.....	19
2.3 Bilddatenverarbeitung.....	19
2.4 Training der KI-basierten Modelle	20

2.5	<i>Testparameter</i>	22
2.6	<i>Statistik</i>	23
2.7	<i>Ethikvotum</i>	23
3	Ergebnisse – Resultate mit graphischen Darstellungen	24
3.1	<i>Ergebnisse CNN Modell EfficientNet B4</i>	24
3.2	<i>Ergebnisse CNN Modell YOLOv5</i>	27
4	Diskussion	43

Abkürzungen und deren Erklärung

LKH – Landesklinikum

AI – Artificial Intelligence

CNN - Convolutional Neuronal Networks

YOLOv5 – You only look once Version 5

COCO – Common Objects in Context

FN – Falsch negativ

FP- Falsch positiv

TN – Richtig negativ

TP – Richtig positiv

TPR – Richtig positive Rate/Sensitivität

TNR – Richtig negative Rate/Spezifität

PI – Patellainstabilität

PFI – patellofemorale Instabilität

Q-Winkel – Quadrizeps Winkel

MPFL – mediales patellofemorales Ligament

ATTT – anterior Tuberositas Tibiae Transfer

a.p. – anterior posterior

KI – Künstliche Intelligenz

Art. – Articulatio

ID – Identifikations Nummer

Art*. – Arteria

Artt. - Arteriae

V. – Vena

Vv. - Venae

N. – Nervus

Nn. – Nervi

M. – Musculus

Mm. – Musculi

Lig. – Ligamentum

Ligg. - Ligamenta

Cond. – Condylus

sog. – sogenannte

bzw. – beziehungsweise

ca. – circa

Tub. – Tuberculum

Tubb. – Tubercula

Men. – Meniscus

Menn. – Menisci

PACS – Picture Archiving and Communication System

TT-TG-Abstand - Tibial Tuberosity Trochlear Groove-Abstand

DDH – Development Dysplasia of the Hip

DICOM - Digital Imaging and Communications in Medicine

JPEG - Joint Photographic Experts Group

ResNet – Residual Neural Network

HITL – Human in the loop

CVAT – Computer Vision Annotation Tool

Abbildungsverzeichnis

<i>Abbildung 1 Dejour Grad A bis D (modifiziert nach Paech et. al (12)</i>	<i>9</i>
<i>Abbildung 2 Dejour A (oben links), Dejour B (oben rechts), Dejour C (unten links) sowie Dejour D (unten rechts).....</i>	<i>10</i>
<i>Abbildung 3 Confusion Matrix YOLOv5 Test mit 5 Vorhersagewerten (Dejour A-D sowie N).....</i>	<i>28</i>
<i>Abbildung 4 F1-Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD</i>	<i>29</i>
<i>Abbildung 5 Precision Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD</i>	<i>30</i>
<i>Abbildung 6 Precision Recall Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD</i>	<i>31</i>
<i>Abbildung 7 Recall-Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD</i>	<i>32</i>
<i>Abbildung 8 Batch 0 Labels DejourA-D sowie Gruppe Control</i>	<i>33</i>
<i>Abbildung 9 Batch 0 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control</i>	<i>34</i>
<i>Abbildung 10 Batch 1 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control</i>	<i>35</i>
<i>Abbildung 11 Batch 1 Labels der Gruppen DejourA-D sowie Control</i>	<i>36</i>
<i>Abbildung 12 Batch2 Labels der Gruppen DejourA-D sowie Control</i>	<i>37</i>
<i>Abbildung 13 Batch2 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control</i>	<i>38</i>
<i>Abbildung 14 Konfusionsmatrix YOLOv5 DejourA-D sowie Gruppe Negative control</i>	<i>39</i>
<i>Abbildung 15 F1-Confidence Kurve für zusammengefasste Gruppen DejourA-D</i>	<i>40</i>
<i>Abbildung 16 Precision Confidence Kurve der Gruppen DejourA-D sowie der Gruppe Negative Control</i>	<i>40</i>
<i>Abbildung 17 Kurvenverläufe für Precision sowie Recall der zusammengefassten Gruppen DejourA-D sowie der Gruppe Negative Control.....</i>	<i>41</i>
<i>Abbildung 18 Kurvenverläufe für Recall sowie Confidence der zusammengefassten Gruppen DejourA-D sowie der Gruppe Negative Control.....</i>	<i>42</i>

Tabellenverzeichnis

<i>Tabelle 11</i>	<i>Testung EfficientNet B4 Dejour Gruppen A-D sowie N.....</i>	<i>25</i>
<i>Tabelle 22</i>	<i>Testung EfficientNet B4 Dejour Gruppen A-D</i>	<i>26</i>
<i>Tabelle 33</i>	<i>Testung EfficientNet B4 Gruppen YES (Dejour A-D) und N (nicht pathologisch)</i>	<i>27</i>

Einleitung - Grundlagen

1.1 Kniegelenk (Art. genu)

1.1.1 Anatomie des Kniegelenks

Das Kniegelenk ist das größte Gelenk des menschlichen Körpers (1-3). Die beiden Hauptaufgaben bestehen in der Aufrechterhaltung der Stabilität durch ossäre und nicht ossäre Strukturen bei gleichzeitiger größtmöglicher Mobilitätsgewährleistung (2, 4).

Das Kniegelenk selbst wird als Art. composita bezeichnet und setzt sich aus drei ossären Komponenten zusammen: (1-3)

- 1.) Femur
- 2.) Tibia
- 3.) Patella

Diese drei ossären Bestandteile tragen jeweils überknorpelte Gelenksflächen, sog. Facies articulares, welche für die Gelenkstätigkeit essenziell sind.

Durch verschiedenste Translations- sowie Rotationsbewegungen kann das Kniegelenk funktionell als Trochogynglimus bezeichnet werden (2, 3). Es besteht in seiner Gesamtheit aus zwei verschiedenen Teilgelenken, der Art. femorotibialis sowie der Art. femoropatellaris. Das Gelenk ist ventral, medial sowie lateral mit Weichteilen bedeckt. Ventral befindet sich das Lig. patellae, welches die Patella mit dem Unterschenkel verbindet. Über die Patella findet die Kraftübertragung des Musculus quadriceps femoris statt. Die Patella sowie das Lig. patellae sind ventral gut tastbar und können als Orientierungspunkte dienen (1, 3, 4).

Dorsal befindet sich die Fossa poplitea, welche einige wichtige Leitungsbahnen wie die Art*. et V. poplitea sowie die Nn. tibialis et peroneus communis beherbergt. Sie wird medial von den Mm. semimembranosus et semitendinosus und dem Caput mediale mm. gastrocnemii begrenzt. Laterale Begrenzung bilden das Caput laterale m. gastrocnemii sowie der M. biceps femoris, welcher nach distal zum Caput fibulae hin an seinen Ansatz führt (1-3).

1.1.2 Knochen, Knorpel und Capsula synovialis

Das Kniegelenk besteht aus drei Knochen. Das Femur trägt die beiden gelenksbildenden Fortsätze *Cond. femoris medialis et lateralis* mit den jeweiligen *Facies articulares*, welche überknorpelt sind. Diese beiden sind ventral durch einen knorpeligen Steg verbunden, welcher mit der *Facies patellaris ossis patellae* artikuliert. Dorsal sind die beiden Kondylen durch die *Fossa intercondylaris* voneinander getrennt. Diese Grube ist nicht mit Knorpel überzogen und wirkt somit nicht an der Gelenkbewegung mit (1, 3).

Gelenksbildend an der Tibia sind der die? *Cond. tibialis medialis et lateralis* mit den beiden ebenfalls überknorpelten *Facies articulares superiores*. Zwischen ihnen liegt die *Eminentia intercondylaris* mit den beiden *Tubb. intercondylaria medialis et lateralis*. Ventral sowie dorsal davon können noch eine *Area intercondylaris anterior et posterior* beschrieben werden (1, 3).

Die Patella stellt das größte Sesambein des menschlichen Körpers dar. Sie besteht aus einer *Facies ventralis et dorsalis*, wobei nur die *Facies dorsalis* mit Gelenkknorpel überzogen ist. Es kann desweiteren eine *Basis patellae*, welche nach proximal weist, sowie ein *Apex patellae*, welcher nach distal weist, beschrieben werden. Die Patella bildet über die Umlenkung der Muskelkraft des *M. quadriceps femoris* über das *Lig. patellae* eine Drehmomenterhöhung durch eine Verlängerung des Hebelarmes (1-4).

Die genannten Gelenkflächen bilden die Teilgelenke der *Art. composita*, die *Art. femorotibialis* sowie die *Art. femoropatellaris*.

Die Gelenkkapsel (*Capsula articularis*) der *Art. genu* wird von 2 verschiedenen Anteilen gebildet. Diese sind die *Capsula synovialis* und die *Capsula fibrosa* (1, 3). Die *Capsula synovialis* beginnt ventral an der Knochen-Knorpel-Grenze des Femurs, wovon aus sie nach proximal zieht, um dort den *Recessus suprapatellaris* zu bilden. Dort befindet sich auch noch eine *Bursa suprapatellaris*, welche mit dem Gelenk kommuniziert. Seitlich und dorsal beginnt die *Capsula synovialis* ebenfalls an der Knorpel-Knochen-Grenze der beiden Kondylen. Tibial wird die Knorpel-Knochen-Grenze als Ansatz der *Membrana synovialis* herangezogen, wobei die *Eminentia intercondylaris* und die *Areae intercondylares ant. et post.* ausgenommen sind. Daher liegen die beiden *Ligg. cruciata*

anterior et posterior zwar innerhalb der Membrana fibrosa aber außerhalb der Membrana synovialis (3).

Die Membrana fibrosa setzt tibial etwa 1cm weiter distal der Knorpel-Knochen Grenze an und umschließt das Kniegelenk zur Gänze, da sie fest mit der Patella und den Seitenbändern verwachsen ist. Femoral umläuft die Membrana fibrosa die Kondylen, ist dorsal an der Linea intercondylaris befestigt und geht ventral von proximal aus in Richtung Quadricepssehne und Patella über in die distalen Anteile (1, 3).

1.1.3 Muskulatur und Weichteilanteile

Ventral am Oberschenkel befindet sich ein prominenter Muskel, dessen Form auch namensgebend ist: der M. quadriceps femoris. Dieser vierköpfige Muskel besteht aus den Mm. vastus lateralis, rectus femoris, vastus medialis und dem vastus intermedius und bildet den Hauptstreckapparat am Kniegelenk (1, 3). Die Streckung ist meist bis 0° aktiv ausführbar und kann passiv meist noch auf etwa 5-10° erhöht werden. Zudem kann durch diesen Muskel die Kniegelenkshöhle gespannt und damit ein Einklemmen der Kapsel bei Streckung im Recessus suprapatellaris verhindert werden (3).

Dorsal befindet sich medial der Fossa poplitea der M. semimembranosus und lateral der M. biceps femoris. Etwas weiter distal finden sich bereits die Ursprünge der beiden Köpfe des M. gastrocnemius, welche die Fossa poplitea nach kaudal hin begrenzen. Desweiteren darf in diesem Bereich der Pes anserinus profundus nicht unerwähnt bleiben, welcher durch die Ansatzsehne des M. semimembranosus im Bereich der proximalen Tibia gebildet wird. Diese teilt sich in fünf verschiedene Zügel an der Insertionsstelle auf und wirkt damit wie ein „Gänsefüßchen“. Funktionell gesehen stabilisiert dieser Anteil das Kniegelenk in leichter Beugung vor der übermäßigen Rotation (1, 3, 4).

Im dorsalen Bereich des Gelenks findet man den M. popliteus, welcher als wichtige Struktur für die Aufhebung der sogenannten Schlussrotation zuständig ist. Die Sehne zieht unter dem Lig. collaterale laterale und einigen Zügeln des Lig. popliteum arcuatum hindurch und in das Gelenk hinein. Sie stellt auch Verbindung her zum lateralen Men. sowie zum Caput fibulae (2-4).

Für die Flexion des Kniegelenks sind mehrere Muskeln zuständig, welche zum Teil bereits zur Unterschenkelmuskulatur zählen. Flexoren sind die Mm. semimembranosus, semitendinosus, biceps femoris, gracilis, popliteus, sartorius sowie die beiden Köpfe des M. gastrocnemius. Funktionell besteht die Flexion aus einer kombinierten Roll- und Gleitbewegung, wobei ein Bewegungsausmaß von 160° passiv und circa 120° (je nach Weichteilsituation) aktiv möglich ist (1-3).

Medialseitig liegt der Pes anserinus superficialis, welcher aus den Mm. sartorius, gracilis und semitendinosus besteht und ebenfalls an der Rotationsstabilität im gebeugten Kniegelenk mitwirkt (3).

An der lateralen Seite kann meist sehr prominent das Tuberculum supracondylare „Gerdy“ getastet werden, an welchem der flächige Tractus iliotibialis ansetzt. Etwas weiter distal befindet sich bereits das Caput fibulae, an welchem noch Ausläufer des Tractus iliotibialis ansetzen können (1, 3).

Befindet sich das Kniegelenk in gebeugter Stellung wird durch die Entspannung der beiden Kollateralbänder eine Rotationskomponente freigegeben. Diese sollte im Gesunden bei 90° gebeugtem Knie bei circa 20° Innenrotation und in etwa 40° Außenrotation liegen. Grund für die geringere Innenrotation ist, dass sich die Kreuzbänder in sich verwirren und somit die Innenrotation stärker limitiert wird. Funktionell gesehen wirken alle im Text oben genannten Muskeln, die medial am Kniegelenk beziehungsweise dem Unterschenkel ansetzen, als Innenrotatoren im Kniegelenk. Der einzige Gegenspieler im Sinne einer Außenrotation ist der M. biceps femoris, welcher an der Fibula ansetzt und im gebeugten Knie die Außenrotation durchführen kann. Rotationsstabilität wird zudem noch durch die beiden Kollateralbänder sowie Kreuzbänder und durch verschiedenste Kapselzüge gewährleistet (1, 3, 4).

Eine Besonderheit hinsichtlich der Bewegungen im Kniegelenk stellt die bereits vorher erwähnte Schlussrotation dar. Diese wird als eine am Ende der Streckung einsetzende Rotation bezeichnet, welche durch die unterschiedliche Form der Femurkondylen, des vorderen Kreuzbandes sowie des Außenbandes zustande kommt. Sie beträgt in etwa 5-10° Außenrotation der Tibia und ermöglicht dadurch den amuskulären Stand (1-3).

Die Aufhebung der Schlussrotation kann durch Anspannung des M. popliteus stattfinden.

Desweiteren befindet sich hinter dem Lig. patellae eine weitere akzessorische Gelenkseinrichtung, ein sog. Corpus adiposum infrapatellare, auch „Hoffa – Fettkörper“ genannt. Dieser ist an der Area intercondylaris anterior befestigt und führt bei Beugung eine Verschiebung nach lateral durch, um für das Ligamentum patellae ausreichend Platz zu schaffen, da dieses nach dorsal gezogen wird (2, 3).

Zusätzlich wird durch den Hoffa Fettkörper die Blutversorgung für das vordere Kreuzband gewährleistet, da aus ihm ein Gefäß für Blut und Nährstoffversorgung des vorderen Kreuzbandes durchtritt. Es können im Kniegelenk noch mehrere Schleimhautfalten, sog. Plicae, beschrieben werden, welche in den meisten Fällen keine klinische Bedeutung haben. Es sind dies die Plicae alares medial sowie lateral der Patella, wobei die mediale Plica bei Hypertrophie zu Schmerzen führen kann (2, 3). Die Plica synovialis medialis gilt als Relikt aus der Embryonalzeit und kann während der Arthroskopie ohne Bedenken entfernt werden (2).

1.1.4 Menisci

Der Men. medialis sowie der Men. lateralis dienen in erster Linie dem Ausgleich der Inkongruenz der beiden Gelenkflächen zwischen dem Tibiaplateau und den Femurkondylen. Des Weiteren können sie Stöße absorbieren, Drücke gleichmäßiger verteilen, unterstützen die Rotation, stabilisieren die Translation und wirken als propriozeptive Sensoren im Kniegelenk (2-5).

Die beiden Menn. bestehen in ihrer Hauptkomponente aus Faserknorpel, welcher mit kollagenen Fasern in unterschiedlichen Faserrichtungen durchzogen ist (1, 3, 5). Diese laufen radiär in Richtung des Mittelpunkts des Meniscus (1).

Die beiden Menn. werden zum Teil durch Diffusion, zum anderen durch Blutgefäße von lateral kommend ernährt, wobei die Versorgung in drei Zonen (von außen zur Gelenksmitte hin) unterteilt werden kann. Die rote Zone mit guter Vaskularisierung, die rot-weiße Zone mit mäßiger bzw. schlechter Vaskularisierung sowie die weiß-weiße Zone ohne Vaskularisierung. Letztere wird ausschließlich durch Diffusion ernährt (2, 5).

Topographisch wird in der Literatur von einem Meniskushinterhorn, einer Intermediärzone und einem Meniskusvorderhorn gesprochen ([2](#), [3](#), [5](#)). Der mediale Men. stellt sich eher halbmondförmig dar und ist mit dem medialen Kollateralband bzw. dessen tiefen Anteilen verwachsen ([1-5](#)). Durch die starre Verankerung an den Areae intercondylares posterior et anterior sowie am medialen Seitenband ist die Beweglichkeit des medialen Men. stärker eingeschränkt als die des Men. lateralis ([1](#), [3](#)). Der laterale Men. ist in seiner Grundform eher O-förmig und nur über Hinter- und Vorderhorn an den Areae intercondylares anterior et posterior verankert. Dies gibt ihm bedeutend mehr Freiraum für Bewegungen, somit ist die Verletzungswahrscheinlichkeit des lateralen Men. um ein Vielfaches geringer einzuschätzen im Vergleich zum Men. medialis ([1-5](#)).

1.1.5 Bandstrukturen

Am Kniegelenk werden Bänder, die an der Gelenksstabilisierung mitwirken sowie Bänder, die Muskelkräfte umlenken, unterschieden. An der Gelenkstabilisierung wirken die beiden Kollateralbänder, das Lig. collaterale mediale et laterale.

Als Muskelkräfte umlenkendes Band ist das Lig. Patellae anzusehen. Es überträgt Kräfte und macht Bewegungen möglich. Dieses Band erstreckt sich vom Apex patellae und inseriert an der Tuberositas tibiae ([1-4](#)).

Zusätzlich können noch verschiedenste sogenannte Retinacula beschrieben werden, welche auch als Reservestreckapparat bezeichnet werden. Dazu gehören das Retinaculum mediale und das Retinaculum laterale ([1-4](#)). Das Retinaculum laterale wird hauptsächlich aus den Fasern des M. vastus lateralis, M. rectus femoris sowie des Tractus iliotibialis gebildet und inseriert lateral der Tuberositas tibiae. Das mediale Retinaculum wird hauptsächlich aus Fasern des M. vastus medialis gebildet und inseriert etwas anterior der Ansatzstelle des Lig. collaterale mediale ([1](#)). beiden Retinacula sichern die Patella zusätzlich vor Luxationen durch Halten der Patella im femoropatellaren Gleitlager ([1-3](#)). Die beiden Vorderhörner der Menn. können durch ein Lig. transversum genus verbunden sein, welches inkonsistent ist ([1-3](#)).

Von größter Bedeutung sind jedoch die beiden Kollateralbänder, welche die laterale sowie mediale Aufklappbarkeit des Kniegelenks verhindern. Das Lig. collaterale laterale

entspringt am Epicondylus lateralis und inseriert an der anterolateralen Tibia, zwischen Tuberositas tibiae und Tub. „Gerdy“. Das mediale Seitenband, Lig. collaterale mediale, entspringt am Epicondylus medialis femoris und zieht breit und ausgedehnt über das Kniegelenk weiter nach distal, um schließlich etwas weiter distal des Tibiakopfplateaus an der Facies medialis tibiae zu inserieren.

Dorsal am Kniegelenk sorgt das Lig. popliteum obliquum gemeinsam mit der Kapselspannung für Stabilität. Es entspringt aus fünf Zügeln des M. semimembranosus und zieht nach proximal und lateral (1, 3). Desweiteren zieht das Lig. popliteum arcuatum, welches vom Apex fibulae ausgeht, von hinten über das Kniegelenk und strahlt über die Sehne des M. popliteus in die Gelenkkapsel ein (1-4). Zwei weitere variabel ausgebildete Bandstrukturen sind die Ligg. meniscofemorales anterior (Humphrey) et posterior (Wrisberg), welche vom lateralen Meniskushinterhorn zur lateralen Seite des medialen Femurkondyls ziehen (1-4).

Die beiden Kreuzbänder, bestehend aus dem Lig. cruciatum anterius und dem Lig. cruciatum posterius sichern das Kniegelenk bei Rotationsbewegungen wie auch bei der anterioposterioren Verschieblichkeit der Tibia gegenüber des Femurs. Sollte eine Instabilität vorhanden sein, wird diese als „Schubladenphänomen“ bezeichnet (3). Desweiteren werden den Ligg. cruciata besondere propriozeptive Funktion zugesprochen, da dort wohl einige Mechanorezeptoren und freie Nervenendigungen sitzen sollen (2).

Das vordere Kreuzband entspringt im Bereich der Innenfläche des Condylus lateralis femoris hinten und zieht an die Area intercondylaris anterior tibiae (1-4). Das hintere Kreuzband entspringt am Condylus medialis tibiae und zieht an die Area intercondylaris posterior tibiae sowie zum Teil noch an die proximale Tibiarückfläche hin (1-3).

1.2 Trochleadysplasie

1.2.1 Definition

Die Trochleadysplasie stellt eine anatomische Anlagestörung dar, welche sich durch eine abnormale Form, der an der Führung der Patella beteiligten Gelenksstrukturen des distalen Femurs, präsentiert. Diese Strukturen, allen voran der sogenannte Sulcus trochlearis femoris, in dem die Patella ihr Gleitlager am Oberschenkelknochen findet, können pathologisch verändert sein und damit eine Luxation der Patella begünstigen. Die Trochleadysplasie stellt damit einen der häufigsten statischen Gründe für eine Patellaluxation dar. Weitere Risikofaktoren sind beispielsweise die Patella alta, Rotationsanomalien, eine valgische Beinachse oder ein außerhalb der Norm befindlicher TT-TG Abstand. Der Tibial Tuberosity Trochlear Groove Abstand entspricht dem Abstand zwischen zwei Messlinien, welche durch die Tuberositas tibiae sowie durch den tiefsten Punkt des Sulcus trochlearis verlaufen und im CT parallel zu den dorsalen Femurkondylen gemessen wird ([6-8](#)).

Die pathologische Anlage der Trochlea femoris kann sowohl die die Tiefe wie auch die grundlegende Form betreffen ([6](#), [9](#)).

Im klinischen Alltag existieren verschiedene Klassifikationen, wobei die Klassifikation nach Dejour in klassischen Röntgenaufnahmen immer noch breite Anwendung findet ([10](#)). Diese beschreibt vier Grade anhand von verschiedenen röntgenologisch feststellbaren pathologischen Parametern: Die Schweregrade sind mit Buchstaben bezeichnet und in ihrer Schwere von A-D aufsteigend: ([6](#), [11](#))

Grundsätzlich werden zur Einteilung der Schwere einer Trochleadysplasie nach Dejour Röntgenaufnahmen in sagittaler Betrachtungsebene gewählt, welche den seitlichen Blick auf das distale Femur zulassen und damit eine Beurteilung der Form der Trochlea femoris ermöglichen. Die Durchführung der Röntgenaufnahme erfolgt in leicht gebeugter Stellung des Kniegelenkes sowie in streng seitlichem Strahlengang, da es für eine adäquate und aussagekräftige Beurteilung wichtig ist, dass sich die beiden Femurkondylen übereinander projizieren ([11](#), [12](#)).

Die unterschiedlichen Schweregrade nach Dejour sind wie folgt charakterisiert (s. auch Abb.1):

Dejour Grad A: Dieser leichtere Grad wird durch eine positives „Crossing-Sign“ (siehe Pfeil Orange) definiert. Das „Crossing-Sign“ selbst wird als ein Überkreuzen der Trochleakontur nach ventral über eine gedachte verlängerte Linie der ventralen Kortikalis des Femurs beschrieben.

Dejour Grad B: In diesem Fall handelt es sich bereits um eine schwerere Dysplasie, welche definiert ist durch einen zusätzlich auftretenden Knochensporn, der sich etwas proximal des Beginns der Trochlea femoris befindet (Pfeile in Schwarz).

Dejour Grad C: Dabei lässt sich eine Doppelkontur des lateralen Kondyls feststellen, welche durch eine Hyperkonvexität des lateralen Kondyls und eine Hypoplasie des medialen Kondyls zustande kommt (Pfeil Grün).

Dejour Grad D: Dieser stellt eine Kombination aus supratrochlearem Sporn und einer Doppelkontur in den lateralen Röntgenaufnahmen dar (Pfeile in Rot) ([9](#), [12-14](#)).

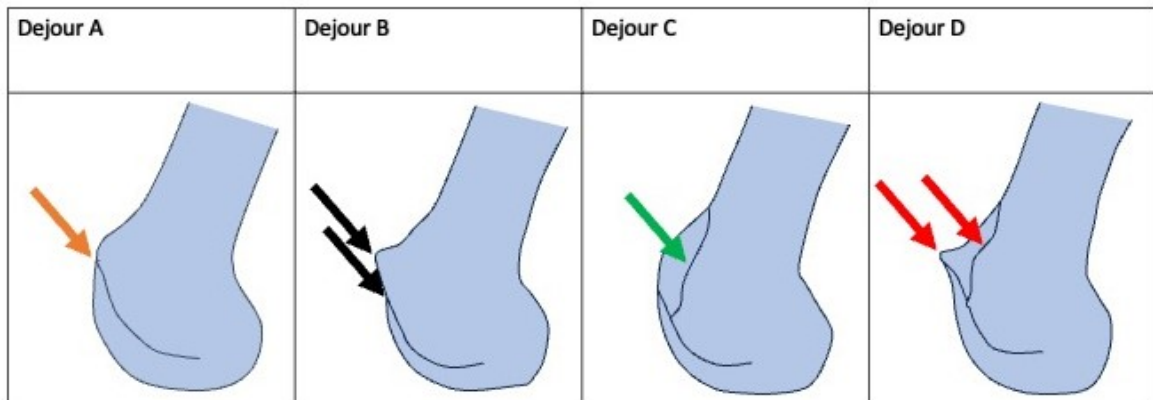


Abbildung 1 Dejour Grad A bis D (modifiziert nach Paech et. al (12))



Abbildung 2 Dejour A (oben links), Dejour B (oben rechts), Dejour C (unten links) sowie Dejour D (unten rechts)

In Abbildung 2 sind beispielhaft je ein Röntgenbild für die jeweiligen Gruppen Dejour A-D aus der für die Studie verwendeten Datenbank dargestellt.

1.2.2 Epidemiologie der Trochleadysplasie

Die Patellaluxation hat eine Prävalenz in der Allgemeinbevölkerung von circa 6/100.000 Personen, mit einem Peak bei den 10–17-Jährigen auf ca. 29/100.000 Personen (15).

Zudem stellt die Patellaluxation die häufigste Ursache einer akuten Knieverletzung im Wachstumsalter dar und ist der häufigste Grund für ein Hämarthros im Wachstumsalter (16). Die Trochleadysplasie gilt als ein prädisponierender Faktor für eine Patellainstabilität. Dejour et al führte bereits 1994 eine anatomisch radiologische Studie mit Patient*innen durch, die unter Patellainstabilität litten, und stellte dabei fest, dass bei 85% der Patient*innen mit Patellainstabilität eine Dysplasie der Trochlea femoris vorlag (6).

Die Trochleadysplasie ist außerdem mit einer höheren Wahrscheinlichkeit für weitere Dysplasien, wie beispielsweise der Hüftdysplasie (DDH), vergesellschaftet. So wurde unter Neugeborenen eine 2,4fach erhöhte Wahrscheinlichkeit für das Vorliegen einer Trochleadysplasie vorgefunden, sofern die Patient*innen unter DDH litten (17).

In einer 2021 durchgeführten epidemiologischen Studie mit 359 untersuchten Kniegelenks-Röntgen wurden in 61.5% der untersuchten Fälle keine pathologische Form, in 21.4% eine milde Form, in 12.7% eine moderate Form sowie in 4.4% eine schwere Form von Trochleadysplasie festgestellt. Es wurden allerdings keine geschlechtsbezogenen Unterschiede hinsichtlich der Schwere der Trochleadysplasie festgestellt (14).

1.2.3 Diagnostik

Patient*innen mit Patellainstabilität werden klinisch häufig vorstellig mit Knieschmerzen, vor allem ventral über der Patella, wohingegen Patient*innen häufig nach einer Patellaluxation oft notfallmäßig vorstellig werden.

Diagnostisch sind folgende Punkte relevant:

- 1.) Anamnese
- 2.) die klinische Diagnostik (im Liegen und im Stehen)
- 3.) die radiologische Diagnostik

Zu Beginn ist eine Anamnese und falls möglich, eine genaue Anamnese des Unfallhergangs sowie eine strukturierte klinische Untersuchung indiziert (18). Die beiden unteren Extremitäten sollten im Seitenvergleich inspiziert werden. Dabei sollten, falls vorhanden, Beinlängendifferenzen, Achsdeviationen oder auch mögliche Muskelasymmetrien diagnostiziert werden. Eine Untersuchung des Bewegungsumfangs des Kniegelenkes passiv sowie aktiv nach der Neutral-Null-Methode sollte jedenfalls durchgeführt werden (19).

Untersuchungen im Liegen:

Im Liegen können der Apprehensions-Test und der Relokationstest nach Fulkerson durchgeführt werden. Beim Apprehensions-Test wird versucht, die Patella nach lateral aus ihrem Gleitlager herauszudrücken/schieben. Wichtig ist, dass die Patient*innen den Musculus quadrizeps entspannen. Während der Ausführung kommt es bei positivem Test zum reflektorischen Anspannen des Muskels, um eine Subluxation/Luxation zu verhindern. Dies gilt als pathognomonisch für eine PFI (18, 20). Beim sogenannten Relokationstest nach Fulkerson wird das Bein vom Untersuchenden gestreckt gehalten, die Patella wird versucht mittels Druckausübung nach medial zu luxieren und währenddessen passiv flektiert. Kommt es zur Subluxation/Luxation oder zu einem spontanen Einhaken der Patella bei Flexion so ist dieser Test positiv und deutet auf eine mediale Instabilität hin (18).

Messung des Q-Winkels:

Es erfolgt die Messung des sogenannten Q-Winkels, welcher sich zwischen zwei Geraden aufspannt: einer Geraden über die Spina iliaca anterior superior hin zum Patellazentrum und einer zweiten Geraden vom Patellazentrum über die Tuberositas tibiae hinweg (21).

Es herrscht jedoch in der Literatur keine absolute Einigkeit, in welcher genauen Stellung sich die Patient*innen während dieser Messung befinden sollten und daher kommt es zu unterschiedlichen Winkel-Gradangaben, welche je nach Kniegelenksstellung unterschiedlich zu interpretieren sind. Um eine valide Messung zu erhalten, sollte die Patella auf jeden Fall zentriert sein und das Bein in neutraler Rotationsstellung gehalten werden (18). Mizuno et al. spricht von einem erhöhten Risiko für laterale Subluxation oder

Luxation bei großem Q-Winkel (22). Pagenstert et al. kamen in ihrer Literaturrecherche zu dem Schluss, dass ein Winkel von $< 20^\circ$ keine Pathologie ausschließt, jedoch ein Winkel $> 20^\circ$ ein Problem im patellofemorale Gelenk darstellen kann, da in diesem Fall von einer, im Sinne der Biomechanik, ungünstigen Gelenksstellung gesprochen werden kann (18).

Untersuchungen im Stehen:

Im Stehen erfolgt die Begutachtung der Beinachse, da eine valgische Achse mit einem erhöhten Aufkommen von patellofemorale Problemen einhergehen kann (19, 23). Bei der Untersuchung im Stehen sollte desweiteren eine Untersuchung von ventral, dorsal sowie seitlich durchgeführt werden, wobei besonders auf die Beinlänge sowie Beinachse geachtet werden muss. Anschließend soll eine Untersuchung des Gangbilds stattfinden. Dabei sollte auf mögliche Auffälligkeiten wie Schonhinken, Ausweich- oder Ersatzbewegungen geachtet werden (19).

Untersuchungen im Sitzen/beim Aufstehen:

Im Sitzen kann auf eine Atrophie des Musculus vastus medialis obliquus geachtet werden. Sollte dies der Fall sein, kann von einer vergrößerten Subluxationswahrscheinlichkeit ausgegangen werden, da die Stabilisierung nach medial damit weniger stark vorhanden ist (18). Desweiteren kann auf die Gleitbewegung der Patella beim Aufstehen geachtet werden. Dieser Test wird als dynamischer Patellatest bezeichnet. Dazu werden die Patient*innen aufgefordert, aus dem Sitzen langsam aufzustehen und die Knie zu strecken oder die Knie im Sitzen zu strecken. Die Patella wird dabei von ventral betrachtet und sollte dabei eine Gerade bestreiten. Liegt ein sogenanntes J-Sign vor, beginnt die Patella mit einer geradlinigen Bewegung und weicht am Ende der Extension nach lateral ab, sie beschreibt also eine umgekehrte J-Form. Dies ist hinweisend auf eine Pathologie im Bereich der Trochlea femoris (18).

Radiologische Diagnostik:

Radiologisch kann eine Diagnostik mittels Röntgen, Magnetresonanztomographie oder auch Computertomographie erfolgen. Die einzelnen diagnostischen Verfahren haben

unterschiedliche Vor- und Nachteile, wobei die Röntgenuntersuchung, die am besten verfügbare und am schnellsten durchführbare Diagnosemöglichkeit darstellt.

Die Röntgenuntersuchung ist zudem kostengünstiger im Vergleich zur CT - oder auch MRT-Untersuchung (24). Im Zuge der Diagnostik der Trochleadysplasie kann das Röntgen als Basisdiagnostik herangezogen werden, wobei vor einem operativen Eingriff ergänzend eine Schnittbilddiagnostik durchgeführt werden sollte (11). Beim Röntgen sollte eine Aufnahme des Kniegelenks in 3 Ebenen stattfinden. Die Standardaufnahmen bestehen aus einem lateralen Strahlengang, einem a.p. Strahlengang sowie einer axialen Patella-Zielaufnahme (11, 12). Bei Verdacht auf Patellaluxation empfiehlt sich eine zusätzliche Patella-Defilée Aufnahme, um einen Eindruck der Patella in ihrem Gleitlager in unterschiedlichen Kniegelenksstellungen zu erhalten. Dazu müssen die Beine in 30°, in 60° sowie in 90° Beugung gelagert werden. In jeder Beugstellung wird jeweils eine axiale Röntgenbildaufnahme angefertigt (11, 12, 25).

Zudem kann eine Ganzbeinstandaufnahme Aufschluss über eine mögliche Achsenfehlstellung geben. Diese kann in Anlehnung an den klinischen Befund erfolgen (23). Die Hauptvorteile der Schichtbilddarstellung wie mit einer MRT sind, die Beurteilung von Weichteilen wie beispielsweise dem medialen patellofemorale Ligament (MPFL) und deren Läsionsmöglichkeiten sowie eine genauere Beurteilung der Morphologie der Trochlea femoris. Auch gibt es bei einer MRT-Untersuchung keine Strahlenbelastungen, was besonders bei Kindern und Jugendlichen nicht außer Acht gelassen werden darf (23, 24).

Messung des TT-TG Abstandes:

Die CT-Untersuchung hat wiederum Vorteile bei Bestimmungen von Torsionsfehlern oder genaueren Messungen wie dem TT-TG-Abstand. Dies ist ein Maß für die Längendifferenz zwischen der Tuberositas tibiae und der „trochlear groove“ (tiefster Punkt der Trochlea femoris) in Bezug zur Referenzlinie (Verbindungsline zwischen den beiden femoralen Kondylenhinterkanten). Dieser wird je nach Literatur bis <14/15mm beziehungsweise <20mm als normal, zwischen 15 und 20mm wird eine OP-Indikation kontrovers diskutiert, ab >20mm als absolut pathologisch eingestuft (12, 23, 26).

1.2.4 Therapie

Im Folgenden soll hauptsächlich auf die Therapie der Trochleadysplasie eingegangen werden. Die Trochleadysplasie gilt als wesentlicher knöcherner Risikofaktor für Patellaluxationen. In den meisten Fällen muss aber auf weitere vorliegende Pathologien, die jeweilige individuelle Anatomie sowie auch das Patient*innenalter eingegangen werden, um die Patient*innen adäquat zu behandeln (27).

Geht man von einer erstmaligen Patellaluxation aus, kann diese, wenn keine osteochondralen Frakturen („flake fractures“) vorliegen, auch konservativ behandelt werden (28). In der Literatur herrscht aber Uneinigkeit darüber, ob nicht bereits die Erstluxation auch schon einer operativen Therapie zugeführt werden sollte. So kommen einige Studien zum Ergebnis, dass die Reluxationsrate beziehungsweise die Instabilität bei primär konservativen Ansätzen erhöht sei (29-32). Bei einer Patellaluxation mit Trochleadysplasie können diverse operative Verfahren zum Einsatz kommen, wobei bei Kindern zumeist nur Weichteileingriffe vorgenommen werden und von knöchernen Transpositionen Abstand genommen wird (33). Die Trochleoplastik stellt beim Erwachsenen die häufigste Methode zur Korrektur einer Trochleadysplasie mit Patellaluxation dar. Sie wird oft gemeinsam mit anderen operativen Verfahren wie der MPFL-Rekonstruktionen oder dem ATTT (anterior tuberositas tibiae transfer) kombiniert, um alle prädisponierenden Faktoren zu adressieren (27). Wurde eine MPFL-Rekonstruktion als alleinige Operation bei Patellaluxation mit Trochleadysplasie durchgeführt, zeigte sich, dass die Rate an Rezidiven und somit auch die Rate an Revisionsoperationen höher war (34). Deshalb befürworten einige Studie im Falle von Dejour B, C und Dejour D Trochleadysplasien zusätzlich eine Trochleoplastik anzustreben, um die (7) Rate an Reluxationen zu minimieren beziehungsweise sogar auf nahezu 0% zu reduzieren (35, 36).

Abschließend kann also festgehalten werden, dass die Therapieentscheidung im Falle einer PFI beziehungsweise Patellaluxation mit Trochleadysplasie von vielen verschiedenen Faktoren abhängig gemacht werden muss, um für Patient*innen ein bestmögliches Therapieergebnis zu erhalten.

1.2.5 Convolutional Neural Network (CNN)

CNN-Netzwerke stellen eine Untergruppe aus dem großem Bereich der Artificial Intelligence dar. Diese Netzwerke werden in der Medizin seit geraumer Zeit eingesetzt, um verschiedenste Hilfestellungen zu geben.

CNN-Netzwerke bestehen aus verschiedenen Blöcken bzw. Schichten („layers“), welche in sogenannte convolution layers, pooling layers sowie fully connected layers eingeteilt werden. Diese drei Blöcke besitzen unterschiedliche Aufgabenverteilungen, wobei im Bereich der convolution layers sowie pooling layers eine Extraktion eines Merkmals stattfindet, welches im Bereich der fully connected layers in eine finale Form, beispielsweise eine Klassifikation, gebracht wird. Das Netzwerk selbst lernt durch Training über eine vorher angefertigte Datenbank, wobei im Hintergrund verschiedenste mathematische Operationen durchgeführt werden. Diese analysieren Gittermuster, wie sie in Pixeln einer Bilddatei vorkommen, weshalb die Netzwerke sehr effizient in der Verarbeitung von Bilddaten sind. Jede Schicht gibt ihre gewonnenen Daten an die nächste Schicht weiter, wodurch der gesamte Prozess immer komplexer wird. Innerhalb des Trainings existieren verschiedene Mechanismen, welche dafür sorgen, dass das Netzwerk einen Abgleich zwischen der selbst errechneten „Vorhersage“ und der in der Datenbank hinterlegten wahren Aussage herstellen. Anhand dieses Abgleichs werden innerhalb der Trainingsphase vom Netzwerk selbst Updates vorgenommen, um die bestmöglichen Ergebnisse zu erzielen. Daher lernt das Netzwerk iterativ und entwickelt sich selbst weiter ([37](#), [38](#)).

Ein sehr komplexes CNN-Modell wie „VGG16“ wurde 2014 veröffentlicht und bietet im Bereich der Bilderkennung mit großen Datensätzen immer noch gute Ergebnisse und wird häufig eingesetzt. Das Training dieses Netzwerkes ist aber aufgrund der großen Parameteranzahl zeitaufwändiger im Vergleich zu den neueren Netzwerken. Deshalb kommen seit 2015 sogenannte „ResNet-CNN“ zum Einsatz. Diese Netzwerke bilden eine Zusammenfassung von verschiedenen Netzwerken für verschiedene Anwendungszwecke und bieten im Vergleich zu „VGG16“ eine geringere Komplexität. Dadurch konnten beispielsweise mit EfficientNet, welches 2019 veröffentlicht wurde, genauere Ergebnisse bei kürzeren Trainingszeiten erzielt, zusätzlich die Genauigkeit durch eine

zusammengesetzte Skalierungsmethode für Tiefe, Breite und Auflösung sogar noch weiter erhöht werden (EfficientNet-B0-EfficientNet-B7) ([39-41](#)).

In der Medizin wurden bereits mehrfach positive Ergebnisse mithilfe von CNNs erzielt.

Beispielsweise wurde Lungentuberkulose auf Thoraxröntgenbildern erkannt ([42](#)).

Desweiteren wurden anhand von Röntgenbildern anatomische Regionen klassifiziert ([43](#)).

Ziel der vorliegenden Arbeit war es, ein AI-Modell hinsichtlich femoraler

Trochleadysplasie zu trainieren, um eine valide KI-generierte Vorhersage zu treffen, ob

das präsentierte Knie Röntgenbild entsprechend der Dejour Klassifikation als pathologisch

zu werten ist oder nicht.

2 Material und Methoden

In diese Studie wurden Patient*innen eingeschlossen, welche zwischen den Jahren 1995

und 2023 mit der Diagnose „Patellaluxation“ am Kinderzentrum der Medizinischen

Universität Graz vorstellig waren. Insgesamt wurden 769 Patient*innen in diese Studie

inkludiert. Bei 650 Patient*innen lag ein Kniegelenksröntgen in 3 Ebenen vor. Nach

Anwendung der Ausschlusskriterien wurden 457 Röntgenbilder für diese Studie

verwendet. Die Ausschlusskriterien lauteten wie folgt:

- 1.) stattgehabte Operation am Kniegelenk/ resp. der Trochlea femoris
- 2.) starke Abweichung von exakt lateraler Projektion des Röntgenbildes (in etwa 15°)

Die ausgehobenen Kniegelenksröntgen wurden 2523 Knie Röntgenbildern von Kindern und Jugendlichen gegenübergestellt, welche im Rahmen von Kontrolluntersuchungen an der klinischen Abteilung für Kinderradiologie des Universitätsklinikum Graz durchgeführt wurden und in einer Datenbank der Kinderradiologie bereits vorhanden waren.

Die Kontrollröntgenbilder wie auch die Röntgenbilder der Patient*innen mit der Diagnose „Patellaluxation“ wurden anonymisiert an einen lokalen Computer der Kinderradiologie transferiert.

2.1 Datenanonymisierung

Im Anschluss an die Datensatzerstellung wurden die Röntgenbilder anonymisiert. Hierzu wurde jedes einzelne Bild über ein speziell programmiertes Python Skript mit einer individuellen ID versehen. Als Basis diente ein Paket mit dem Namen „pydicom“ (44).

Im ersten Schritt wurden die Dateinamen anhand der DICOM-Header Informationen erstellt. Mit Hilfe von „blake2b“ (45) wurden kryptographische Hashes bestehend aus „Institution“ und „Patient*innen ID“ herangezogen, um verkettete Zeichenfolgen zu erhalten. Um die schlussendlichen Zeichenfolgen zu erhalten, wurden die oben erstellten mit „SHA-3-256“ nochmals gekreuzt (46). Durch diese Prozedur erhielt jede*r Patient*in einen individuellen Schlüssel aus Zahlen und Buchstaben, der aber nicht zurückverfolgbar ist (rot markierter Teil der ID).

Der nachfolgende Teil (grün markierter Teil der ID) des Dateinamens wurde durch folgende Schritte erstellt: vom Zeitstempel der Erfassungszeit wurde ein 9-stelliger, aus der „Patient*innen ID“ erstellten Hash subtrahiert und damit die originalen Aufnahmedaten irreversibel verschleiert, wobei die Intervalle zwischen den einzelnen Daten aber noch erhalten wurden. Zum dritten Teil (blau markierter Teil der ID) gehören eine Seriennummer sowie eine Bildnummer, welche jeweils aus zwei Ziffern bestehen. Anschließend stehen Abkürzungen für die anatomische Region des Röntgenbilds sowie abschließend Geschlecht (47).

Beispielhaft für die Anonymisierung hier eine ID:

ID_000619_cded392921f893a07ce1cec48b801aa1a95573baad9c34dd30f64e7077737cc4_0385456338_02-02_LEG-N0_F.jpg.

Das Alter der Patient*innen wurde aus den DICOM-Header Informationen extrahiert und in der Datenbank übernommen.

Nach der Erstellung der Datenbank wurden die Bilder manuell gesichtet und vom Verfasser anhand der Dejour Klassifikation in die Gruppen Dejour A, Dejour B, Dejour C sowie Dejour D eingeteilt. Röntgenbilder, welche stark von einer seitlichen Projektion des Knieröntgens abwichen, wurden in diesem Schritt aus der Datenbank entfernt. Somit

konnten schlussendlich die Gruppen Dejour A mit n=97, Dejour B mit n=59, Dejour C mit n=175 sowie Dejour D mit n=126 Röntgenbildern erstellt werden.

Die Vergleichsgruppe ohne Pathologie umfasste nach Durchsicht und Ausschluss von Röntgenbildern mit Frakturen oder anderen knöchernen Pathologien eine Zahl von n=475. Gesamt wurden somit 932 Röntgenbilder in die Arbeit eingeschlossen.

Für die Erstellung eines Testsatzes wurden zufällig 80 Röntgenbilder aus den vier Dejour Gruppen A-D separiert. Aus der Gruppe der als normal definierten Bilder wurden ebenfalls insgesamt 80 Bilder zufällig als Testdatensatz exprimiert. Die übrigen Röntgenbilder in den Dejour Gruppen A-D wurden für das Training eines Convolutional Neural Network (CNN)-Modells bereitgestellt.

2.2 Röntgenbild Selektion

Die in der Studie verwendeten Röntgenbilder wurden an der Universitätsklinik für Radiologie, Klinische Abteilung für Kinderradiologie, Medizinische Universität Graz angefertigt und anschließend im lokalen System abgespeichert. Das lokale Bildarchiv wird Picture Archiving and Communication System (PACS) genannt. Dort sind die Bilddaten im Digital Imaging and Communications in Medicine Format (DICOM) abgespeichert.

2.3 Bilddatenverarbeitung

Die DICOM Bilddateien wurden nach dem Download aus dem PACS auf einem lokalen Netzwerk gespeichert und anschließend in das gängige Bildformat JPEG umgewandelt. Die DICOM Bilddateien bestanden aus verschiedenen Bildgrößen mit unterschiedlichen Graustufen zwischen 12-Bit beziehungsweise 16-Bit. Um die Bilder auf die Bearbeitung mit den beiden CNN-Algorithmen vorzubereiten, wurden folgende unten im Text beschriebenen Schritte durchgeführt.

Die Konvertierung der DICOM Bilddateien in 16-Bit JPEG-Dateien wurde mithilfe eines Python Packages namens „pydicom“ durchgeführt. So wurde die Pixelgrößen im Original belassen, die unterschiedlichen Graustufen hingegen auf ein Spektrum von 16bit

standardisiert. Zusätzlich wurden die Röntgenbilder über ein Post-Processing mittels „exposure“ – Modul des Python „scikit-image“- Package von ursprünglichen 16-bit Input in float64 konvertiert. Bei dieser Konvertierung wurden die Graustufen durch 65,535 geteilt. Die Intensität der JPEG-Dateien wurden dann um die oberen sowie unteren 0,05% Perzentilen der Bild Histogramme beschnitten. Abschließend wurde mittels „exposure adapthist“ über „Standardeinstellungen“ eine Kontrastverbesserung durchgeführt. Die somit erhaltenen float64-Bilddaten wurden in 8-bit konvertiert und abgespeichert. (47)

2.4 Training der KI-basierten Modelle

In dieser Studie wurden zwei verschiedene CNN-Modelle zur Auswertung herangezogen. Diese tragen die Bezeichnungen EfficientNet-B4 sowie YOLOv5 (39, 48). Aufgrund von Einschränkungen in Bezug auf den Graphikspeicher konnte das CNN-Modell EfficientNet-B7 nicht zur Anwendung kommen.

Die oben genannten Modelle wurden allesamt anhand des Imagenet Datensatzes vortrainiert. Anhand dieses Datensatzes konnten für die Studie die Prinzipien des sogenannten Transfer-Lernen angewendet werden. Bei diesem Transfer-Lernen werden die tiefen Ebenen des CNN nicht verändert, die oberflächlicheren Ebenen des CNN aber sehr wohl trainiert. Diese zuletzt genannten „Layer des CNN“ sind für die Vorhersage eines Bildes essenziell. Die Trainingszeiten des CNN können anhand dieser Methoden deutlich verkürzt werden (49).

In der Studie wurden die neuronalen Netzwerke auf einem Personalcomputer mit Linux Betriebssystem und 8 Gigabyte Grafikkartenspeicher trainiert. Zum Einlesen und Trainieren der CNN-Modelle wurde die FastAI Python-Bibliothek verwendet (50). Die Röntgenbilder wurden nach den oben genannten Bildbearbeitungen in das Trainingsmodell eingefügt.

Um die Modelle weniger anfällig für Datenvariabilitäten zu machen, wurde ein FastAI-Modul namens „aug_transformers“ angewendet. Dieses Modul führt folgende Schritte auf jedem Bild durch: zufälliges Umdrehen mit einer Wahrscheinlichkeit von $p=0,5$, zufälliges Drehen des Bildes um maximal 10° mit einer Wahrscheinlichkeit von $p=0,75$, Erhöhung

von Kontrast und Helligkeit um den Faktor 0,2 mit einer Wahrscheinlichkeit von 0,75 und abschließend ein Zoomen in einen zufällig ausgewählten Bereich des Bildes mit einer Zoomrate zwischen 1,0 und 1,1 und einer Wahrscheinlichkeit von 0,75.

Alle vorher genannten Modelle wurden mit 100 Zyklen (= Epochen oder „Epochs“) trainiert, um eine bestmögliche Performance zu erhalten. Anhand des Validierungssatzes konnten die leistungsstärksten Kombinationen aus den Modellen ausgewertet und gespeichert werden. Die besten Validierungswerte erreichten die Modelle in Experimenten um den Zyklus 40, wodurch 100 Zyklen mehr als genug waren, um mit gegebenen Parametern das bestmögliche Ergebnis zu erreichen. Der Optimizer Adam erzielte eine Lernkurve von $\alpha = 0,0001$.

Da nun bereits eine Datenbank mit pathologischen beziehungsweise nach Dejour klassifizierten Knie Röntgenbildern sowie nicht pathologischen Knie Röntgenbildern erstellt worden war, wurde ein weiteres AI-Tool angewandt: eine computerbasierte Objekterkennung. Es existieren bereits einige Studien, die positive Ergebnisse im Umgang mit großen Datenmengen im Zusammenhang mit Erkennung von Pathologien bei Röntgenbildern erzielen konnten ([51-53](#)).

Es wurden 5 Gruppen erstellt: Dejour-Gruppen A, B, C und D mit jeweils pathologischen Röntgenbildern sowie Gruppe N mit nicht pathologischen Röntgenbildern. Die anonymisierten Röntgenbilder wurden über ein online verfügbares Computer Vision Annotation Tool (CVAT) ([54](#)), auf welches mit einem handelsüblichen Webbrowser zugegriffen wurde, hochgeladen, um sie dort mit verschiedenen Labels zu versehen.

So wurden auf die Gruppennamen jeweils gleichlautende Labels erstellt, anhand derer anschließend die Objekterkennung stattfinden sollte.

Beispielsweise wurde für die Gruppe Dejour A ein Label Dejour A, Gruppe Dejour B ein Label Dejour B usw. erstellt.

Mit Hilfe des CVAT konnte der „Point of Interest“, das heißt jener Bereich am Röntgenbild, welcher für die Klassifikation nach Dejour entscheidend ist, markiert werden. Dazu wurde über jedem Röntgenbild ein Rechteck in variabler Größe über dem Point of Interest aufgezogen. Durch das Aufziehen des Rechteckes über dem jeweiligen

Röntgenbild, wurde das Röntgenbild gleichzeitig mit dem jeweils vorausgewählten zugehörigen Label versehen.

Anschließend wurden die Bilder mit einem State-of-the-art Neuronales Netzwerk für Objekterkennung namens YOLOv5 trainiert (55). Dazu wurden aus der erstellten Datenbank der Röntgenbilder 200 Bilder zufällig separiert um als Trainingsset zu fungieren. Die restlichen Röntgenbilder wurden als Kontrollset benutzt. Auf einem Linux basierten Windows PC wurden nun 100 „Epochs“ oder Trainingszyklen mit dem YOLOv5-Modell, welches unter „COCO“ vortrainiert wurde, ausgeführt (56).

2.5 Testparameter

Die weit verbreiteten Parameter für die Performance von Neuronalen Netzwerken wie Sensitivität (TPR), Spezifität (TNR), F1-Score (F1) sowie der Youdens-Index(J) kamen auch in diesem Fall zur Anwendung (57). Um die Kennzahlen für den Test zu ermitteln, wurden Konfusionsmatrizen erstellt. Diese enthalten „richtig positive“ (TP), „falsch positive“ (FP), „richtig negative“ (TN) sowie „falsch negative“ (FN) in einem Raster. Verglichen werden positive sowie negative Vorhersagen mit realen Bedingungen.

Sensitivität (=recall): $TPR = \frac{TP}{TP+FN}$

Spezifität: $TNR = \frac{TN}{TN+FP}$

Positiver Vorhersage Wert (=precision): $PPV = \frac{TP}{TP+FP}$

Negativer Vorhersage Wert: $NPV = \frac{TN}{TN+FN}$

Genauigkeit (=accuracy): $ACC = \frac{TP+TN}{TP+TN+FP+FN}$

F1-Score: $F1 = 2 \frac{TP}{2TP+FP+FN}$

Zudem wurden t-Tests zur Analyse der verschiedenen Datensätze in Bezug auf die Altersverteilung durchgeführt. Die Voraussetzungen zur Verwendung der t-Tests wurden zuvor mittels Q-Q Plot visuell überprüft und mittels Kolmogorov–Smirnov Test rechnerisch bestätigt.

2.6 Statistik

Die statistischen Analysen sowie Berechnungen wurden mit IBM SPSS Statistics Version 21 (IBM, Armonk, New York, United States) durchgeführt.

2.7 Ethikvotum

Für die Datenauswertungen dieser Studie wurde um ein Votum der Ethikkommission der Medizinischen Universität Graz (IRB00002556) angesucht. Auf die Einholung individueller Einverständniserklärungen der Patient*innen und der Sorgeberechtigten konnte aufgrund des retrospektiven Studiendesigns verzichtet werden. Die Referenznummer des positiven Votums lautete 34-119 ex 21/22.

3 Ergebnisse – Resultate mit graphischen Darstellungen

In diesem Abschnitt werden die Ergebnisse der durchgeführten Tests mit den trainierten KI-basierten Modellen angeführt. Im ersten Teil werden die Ergebnisse des CNN-Modells EfficientNet B4, im zweiten Teil die Ergebnisse der Objekterkennungsanalyse mittels YOLOv5 präsentiert.

Das Alter der pathologischen Fälle betrug im Mittel 14,5 +/- 2,2 Jahre. Das Alter in der Kontrollgruppe betrug durchschnittlich 13,9 +/- 2,9 Jahre und der Unterschied zwischen den beiden Gruppen war statistisch signifikant (t-Test $p=0.001$).

Der ungesehene Test-Datensatz (jener Datensatz, der für den abschließenden Test im Anschluss an das Training verwendet wurde) wies ein Alter von 14,3 +/- 2,6 Jahre auf, der Kontrolldatensatz hatte ein durchschnittliches Alter von 14,3 +/- 2,4 Jahren. Der Unterschied für diese insgesamt 160 Bilder war nicht statistisch signifikant (t-Test $p=0.959$).

Die Verteilung der Geschlechter war nahezu identisch, mit 51,5% männlichen und 48,5% weiblichen Fällen.

3.1 Ergebnisse CNN Modell EfficientNet B4

Beginnen möchte ich mit den Ergebnissen aus dem CNN Modells EfficientNet B4. Dabei wurden zwei unterschiedliche Studienmodelle gewählt:

1. korrekte Klassifikation in die 4 Dejour Typen A-D oder N (=nicht pathologisch)
2. korrekte Vorhersage von pathologisch vs. nicht pathologisch (YES/NO)

Bei dem Vorhersagetest konnten folgende Ergebnisse erzielt werden.

Gruppen Dejour A-D/ N	Vorhersagen CNN-Modell EfficientNet B4					Gesamtergebnis
	A	B	C	D	N	
A	11		5	1	3	20
B	5	6		8	1	20
C	1		10	2	7	20
D	2		7	8	3	20
N			1		79	80
Gesamtergebnis	19	6	23	19	93	160

Tabelle 1 Testung EfficientNet B4 Dejour Gruppen A-D sowie N

In obiger Tabelle 1 wurden in einem Testdurchlauf pro Gruppe (Dejour A-D sowie N) jeweils 20 Röntgenbilder separiert und für die Testung verwendet. Korrekt zugeordnete Anzahl an Bildern wurde grün markiert und hervorgehoben.

So konnten in der Gruppe Dejour A 55%, in der Gruppe Dejour B 30%, in der Gruppe Dejour C 50% und in der Gruppe Dejour D 40% korrekt zugeordnet werden.

Gruppenübergreifend (Dejour A-D) konnten 43,75% aller Röntgenbilder korrekt klassifiziert werden.

Die Gruppe N hingegen weist eine hohe korrekte Vorhersagerate mit 98,7% auf. Das bedeutet, dass das CNN eine hohe Vorhersagequalität aufweist, sofern das präsentierte Röntgenbild keine Pathologie besitzt.

	Vorhersagen des CNN Modells EfficientNet B4				
Tatsächlich vorliegende Pathologie	A	B	C	D	N pro Gruppe
A	15	1	4		20
B	5	4	4	7	20
C	1	1	14	4	20
D	2		11	7	20
Gesamtergebnis	23	6	33	18	80

Tabelle 2 Testung EfficientNet B4 Dejour Gruppen A-D

Tabelle 2 zeigt die Verteilung eines weiteren Trainingslaufes, durchgeführt mit dem CNN Modell EfficientNet B4. In diesem Fall wurde ausschließlich auf die korrekte Zuordnung (grün markiert) zu den einzelnen pathologischen Gruppen Dejour A-D geachtet und die Gruppe N (=nicht pathologisch) ausgeschlossen. Wie bereits in Tabelle 1 sind in den Zeilen die jeweilige Anzahl an separierten Röntgenbildern der einzelnen Gruppen eingetragen. So wurden auch in diesem Fall pro Gruppe (A-D) 20 Röntgenbilder für den Test verwendet. Davon wurden in der Gruppe A 15 richtig zugeordnet, 1 Bild der Gruppe B und 4 Bilder der Gruppe C. 4 Bilder wurden in der Gruppe B korrekt zugeordnet, 5 Bilder fälschlicherweise der Gruppe A sowie 4 der Gruppe C und 7 der Gruppe D. In der Gruppe C wurden 14 Röntgenbilder korrekt, jeweils 1 Bild den Gruppen A und B sowie 4 Bilder der Gruppe D zugeordnet. 7 Bilder wurden in Gruppe D richtig, 11 der Gruppe C und 2 Bilder der Gruppe A falsch zugeordnet.

Tatsächlich vorliegende Pathologie	N pro Gruppe	Vorhersagen des CNN-Modells EfficientNet B4	
		N pro Gruppe	
		NO	YES
NO	80	79	1
YES	80	21	59
Gesamtergebnis	160	100	60

Tabelle 3 Testung EfficientNet B4 Gruppen YES (Dejour A-D) und N (nicht pathologisch)

Außerdem wurde in einem weiteren Test eine Unterscheidung hinsichtlich Pathologie vorgenommen. Hierzu wurde eine Einteilung in YES (=vorliegende Pathologie entsprechend einer Dejour Gruppe A-D) und NO (=nicht pathologisch) vorgenommen.

In oben dargestellter Tabelle 3 sind in den Zeilen die jeweilige Anzahl an Röntgenbildern, welche der jeweiligen Gruppe richtig zugeordnet wurden, wiederum grün markiert.

Es wurden pro Gruppe (YES/NO) jeweils 80 Röntgenbilder separiert und diese wiederum mit dem Netzwerk EfficientNet B4 analysiert und eine Vorhersage der Klassifizierung vorgenommen. Dabei konnten in der Gruppe NO (=nicht pathologisch) von 80 Röntgenbildern 79 in die richtige Gruppe NO zugeordnet werden und nur 1 einziges Bild wurde falsch klassifiziert. In der Gruppe YES (pathologisch mit Dejour A-D) wurden aus 80 Röntgenbildern 59 richtig zugeordnet und lediglich 21 falsch der Gruppe NO zugeordnet. Der Unterschied zwischen den Gruppen "YES" und "NO" war statistisch signifikant, $p < 0,001$ (Qui-Quadrat-Test). Das bedeutet, dass Patelladysplasien vom neuronalen Netzwerk übersehen werden können.

3.2 Ergebnisse CNN Modell YOLOv5

Desweiteren wurden Vorhersage Tests mit dem Netzwerk YOLOv5 durchgeführt, welches mittels Objekterkennung eine Zuordnung zu den Gruppen vornehmen sollte.

Hierbei wurden wiederum zwei verschiedenen Varianten gewählt: die Zuordnung der Röntgenbilder in die jeweiligen Gruppen Dejour A-D sowie Gruppe N (normal) und eine zweite Zuordnung in YES/NO Gruppen entsprechend „Pathologisch/nicht pathologisch“.



Abbildung 3 Confusion Matrix YOLOv5 Test mit 5 Vorhersagewerten (Dejour A-D sowie N)

Abbildung Nr. 3 Auf der x-Achse sind die Klassifikationen aufgetragen, in welche die getesteten Röntgenbilder tatsächlich eingeordnet wurden, auf der y-Achse sind es die Vorhersagewerte des durchgeführten Testvorganges. Die Matrix zeigt anhand der in den jeweiligen Kästchen hinterlegten Farbe beziehungsweise deren Deckkraft (niedrige Deckkraft = schlechtere Vorhersagewahrscheinlichkeit, hohe Deckkraft = bessere Vorhersagewahrscheinlichkeit) die Wahrscheinlichkeiten einer korrekten Vorhersage an. So konnten in Gruppe Dejour A 60% korrekt zugeordnet werden, in der Gruppe Dejour B waren es 5%. In der Gruppe Dejour C wurden insgesamt 40% korrekt vorhergesagt und in der Gruppe Dejour D waren es ebenfalls 60%. Sieht man sich die Gruppe „Control“ näher an, welche die nicht pathologischen Röntgenbilder enthält, so wurden dort 97% als nicht pathologisch vorhergesagt. Das bedeutet, dass das Modell mit 97% Wahrscheinlichkeit ein nicht pathologisches Röntgenbild als nicht pathologisch klassifiziert und in nur 3% ein

pathologisches Röntgenbild als nicht pathologisch einstuft. Somit ist eine Spezifität von 97% gegeben.

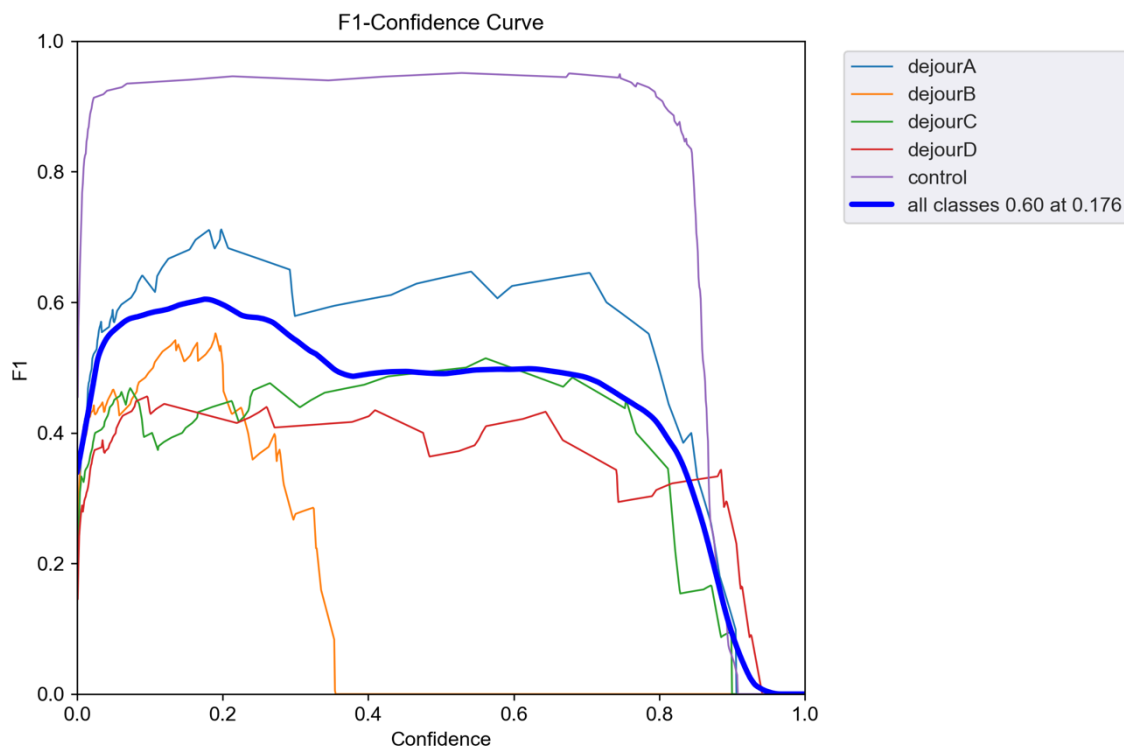


Abbildung 4 F1-Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD

Abbildung 4 zeigt, dass die Gruppe „Control“ sehr gute Vorhersagewahrscheinlichkeiten mit F1 Werten und Confidence Werten von nahezu 1,0 aufweist. Die jeweiligen Dejour Gruppen weisen jedoch geringere Vorhersagewahrscheinlichkeiten in Bezug auf F1- sowie Confidence Werte auf. Die besten Vorhersagewerte bietet die Gruppe Dejour A, wobei die gruppenübergreifende Kurve auf einen maximalen F1 Wert von 0,6 bei einem Confidence Wert von 0,176 kommt.

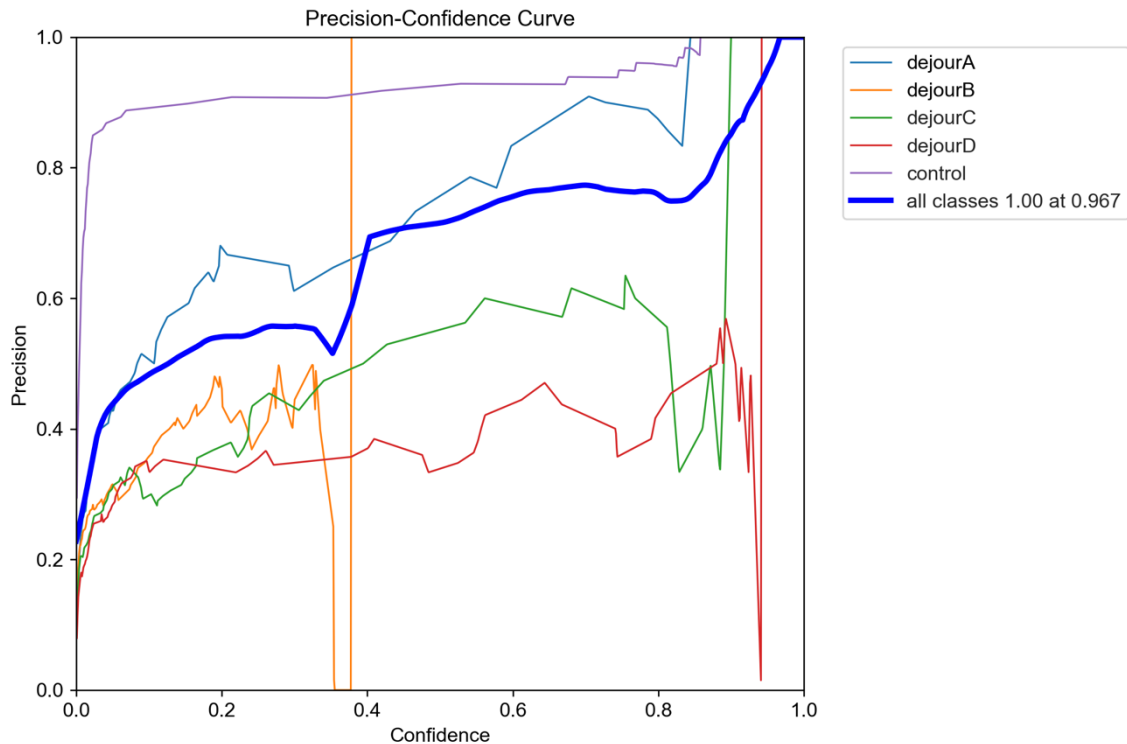


Abbildung 5 Precision Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD

In Abbildung 5 werden die Vorhersagewahrscheinlichkeiten in Bezug auf Precision sowie Confidence veranschaulicht. Dazu wurden auf der x-Achse die Confidence Werte und auf der y-Achse die Precision Werte aufgetragen. Die einzelnen Gruppen von Dejour A-D sowie die Gruppe Control weisen unterschiedliche Kurvenverläufe auf, wobei die Gruppe Control, wie auch in Abbildung 2 dargestellt, die besten Werte aufweist. Die gruppenübergreifende Kurve, in Blau dargestellt, weist eine Precision von 1,00 bei einem Confidence Wert von 0,967 auf. Unter den Gruppen Dejour A-D weist die Gruppe Dejour A die besten Werte auf (türkiser Kurvenverlauf).

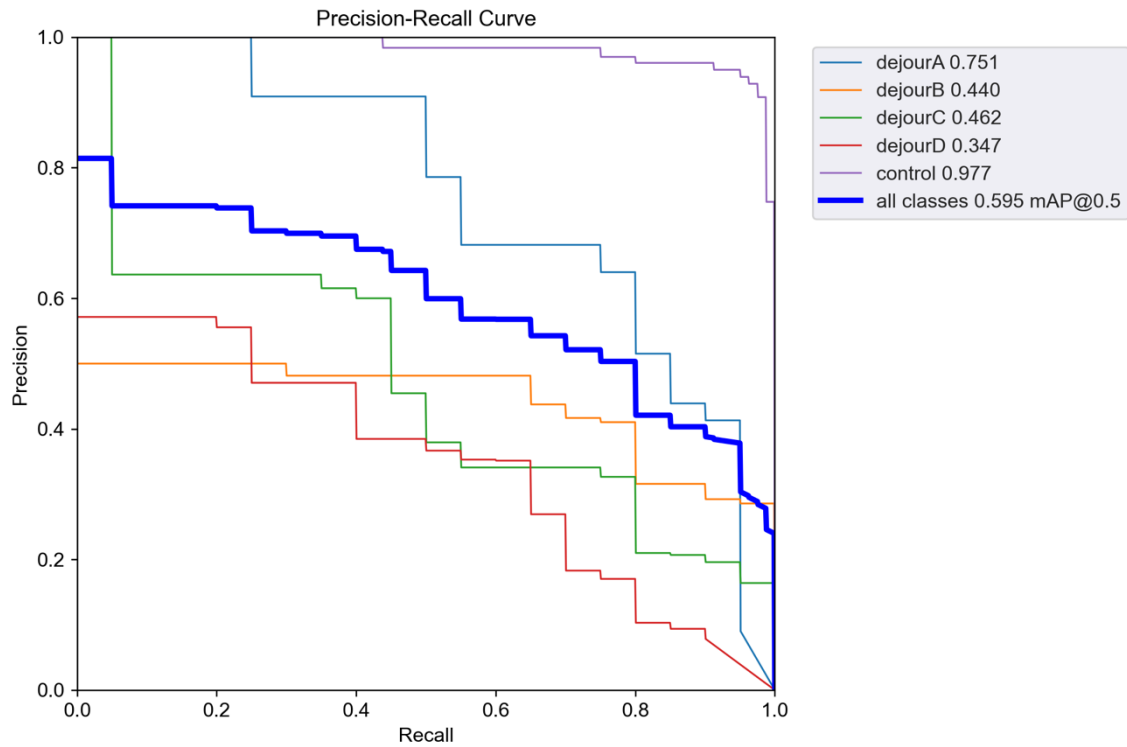


Abbildung 6 Precision Recall Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD

In Abbildung 6 werden Precision zu Recall Kurven miteinander verglichen. Die Recall Werte auf der x-Achse, die y-Achse stellt die Werten der Precision dar. Die höchste Precision unter den Gruppen DejourA-D konnte in der Gruppe Dejour A erreicht werden. Gruppenübergreifend konnte eine maximale Precision von 0,595 bei einem Recall von 0,5 erzielt werden. Die Gruppe Control erreichte die besten Werte mit Precision 0,977.

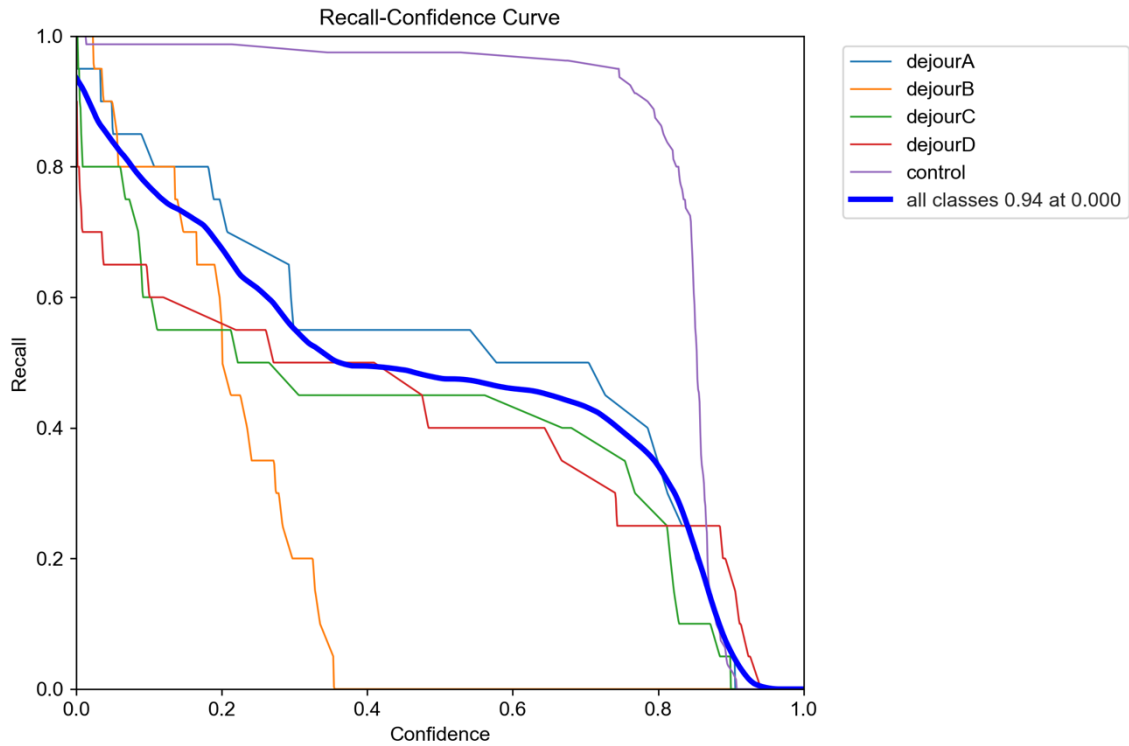


Abbildung 7 Recall-Confidence Kurve der Gruppen DejourA, DejourB, DejourC sowie DejourD

Abbildung 7 zeigt anhand von Kurvenverläufen die Zusammenhänge zwischen Recall auf der x-Achse und Confidence auf der y-Achse der Gruppen Dejour A-D sowie der Gruppe „Control“. Die besten Werte konnten wiederum in der Gruppe Control erzielt werden, wobei in allen Klassen bei 0,00 Confidence ein Wert von 0,94 für den Recall erzielt werden konnte.

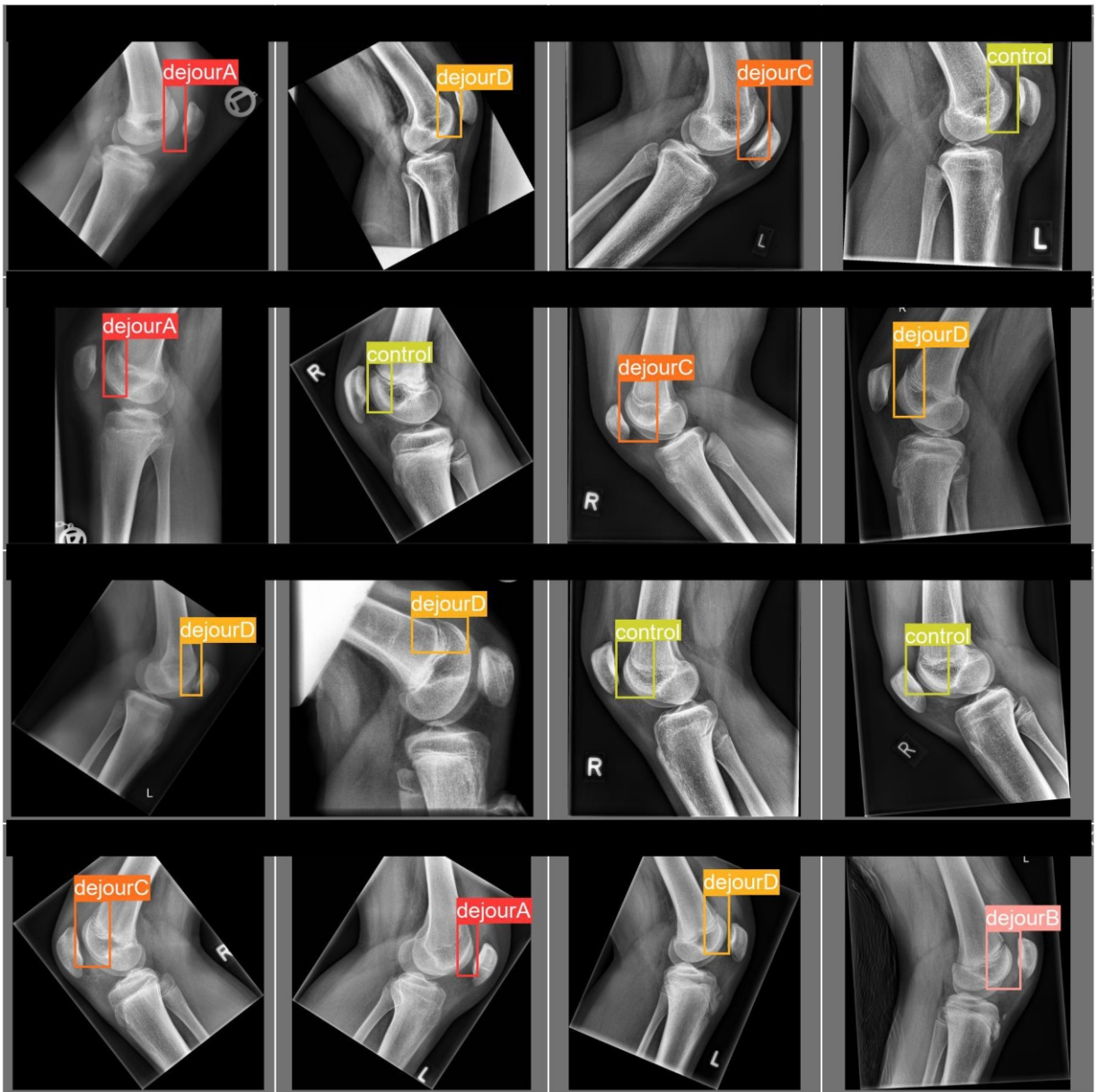


Abbildung 8 Batch 0 Labels DejourA-D sowie Gruppe Control

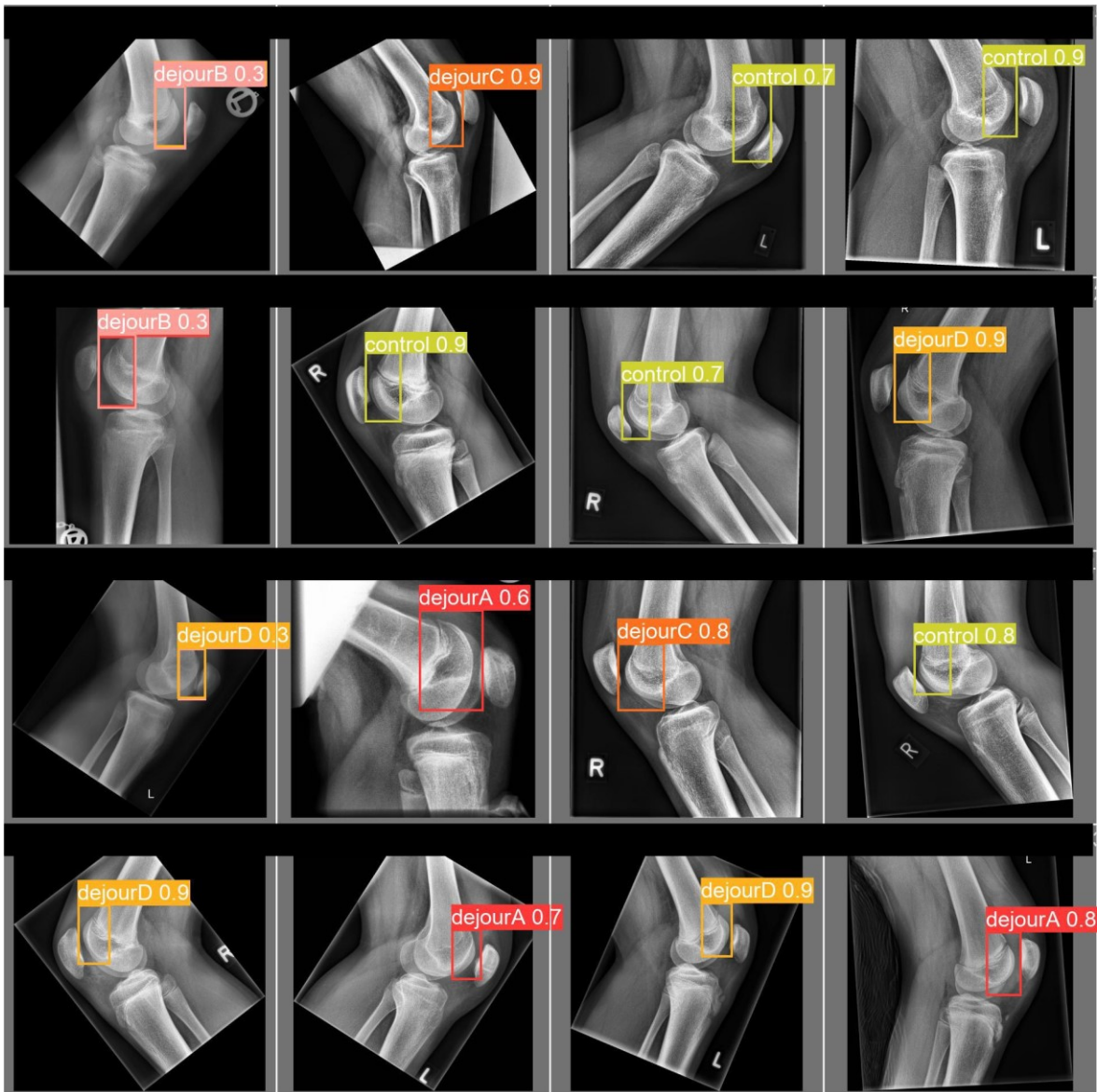


Abbildung 9 Batch 0 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control

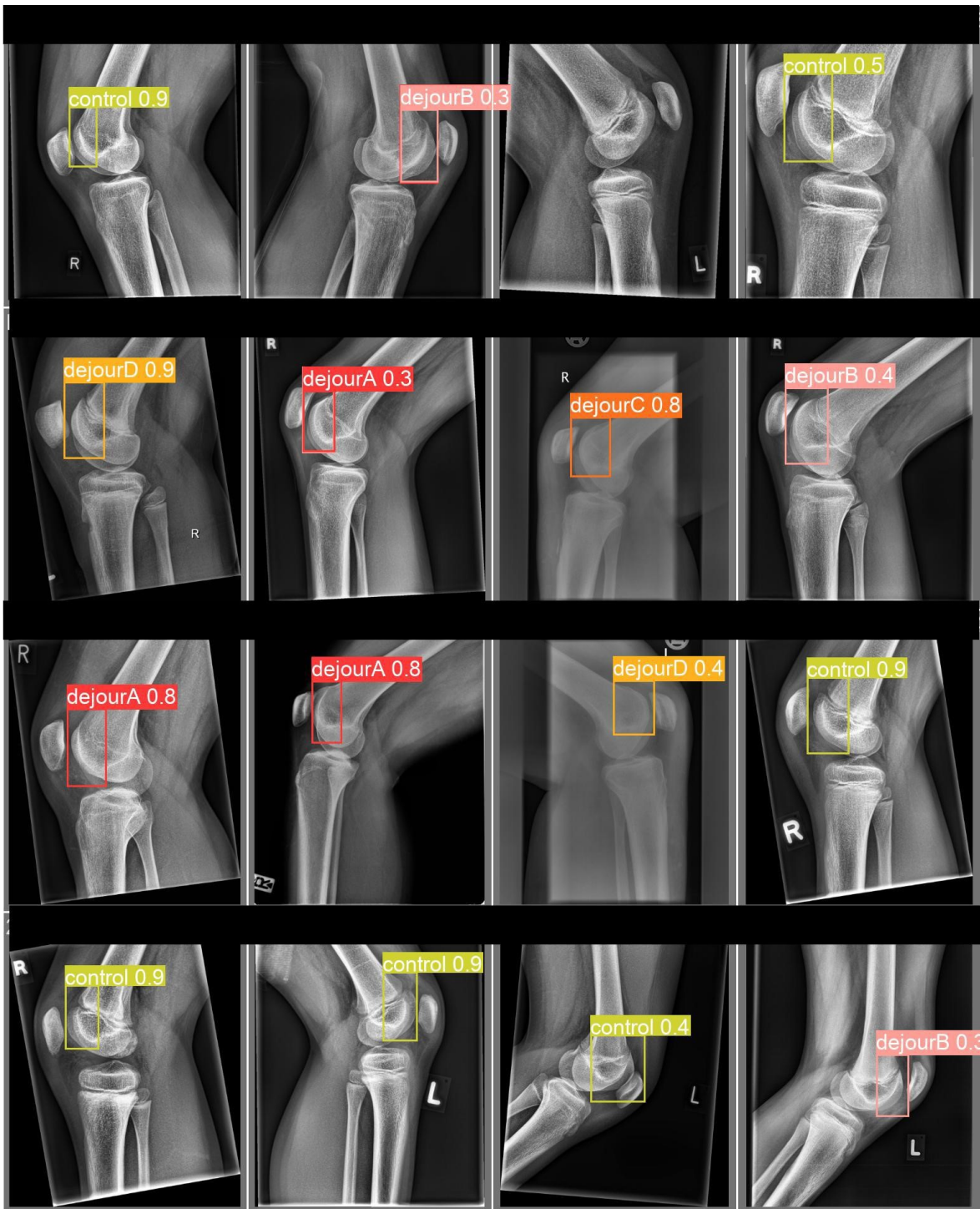


Abbildung 10 Batch 1 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control

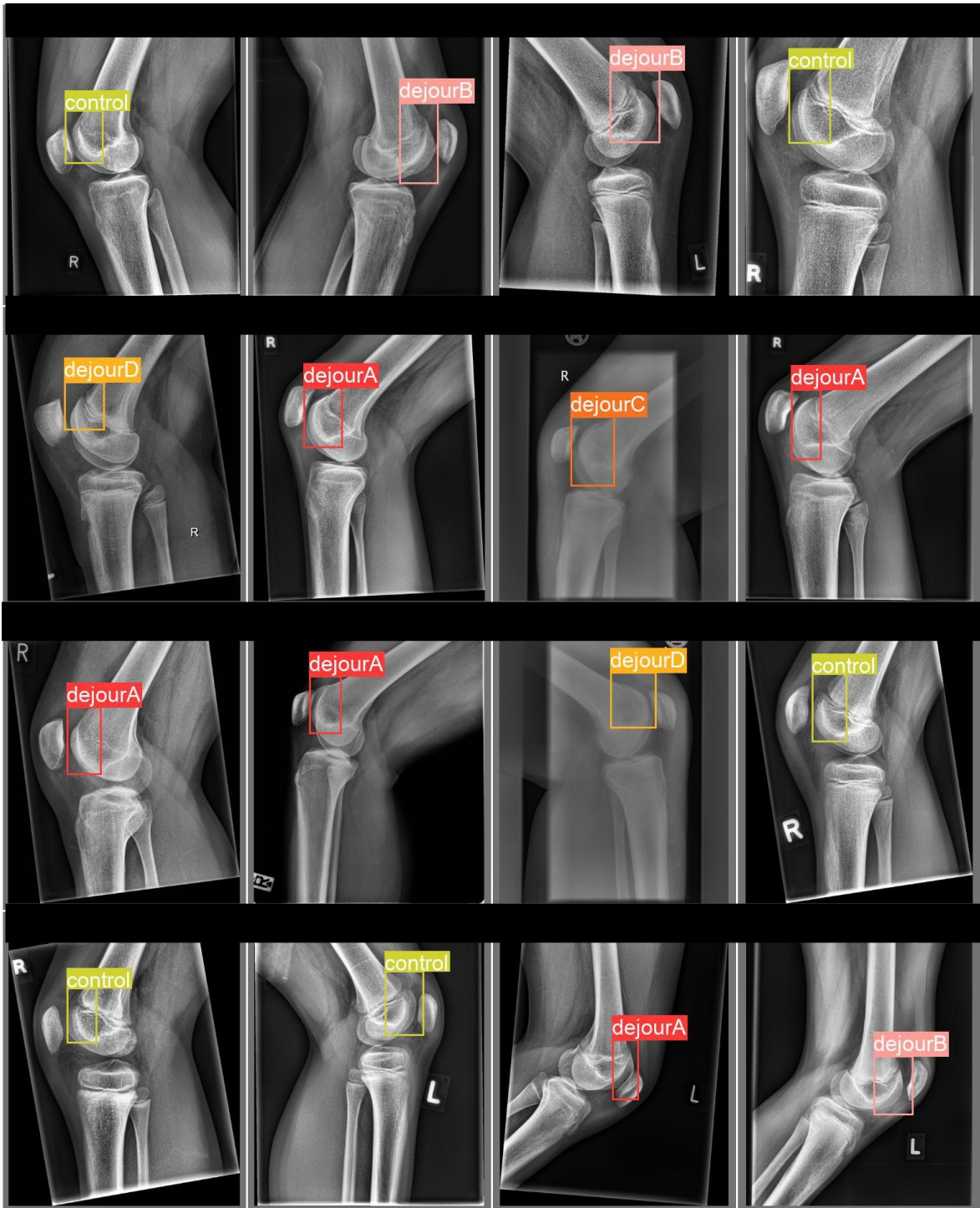


Abbildung 11 Batch 1 Labels der Gruppen DejourA-D sowie Control

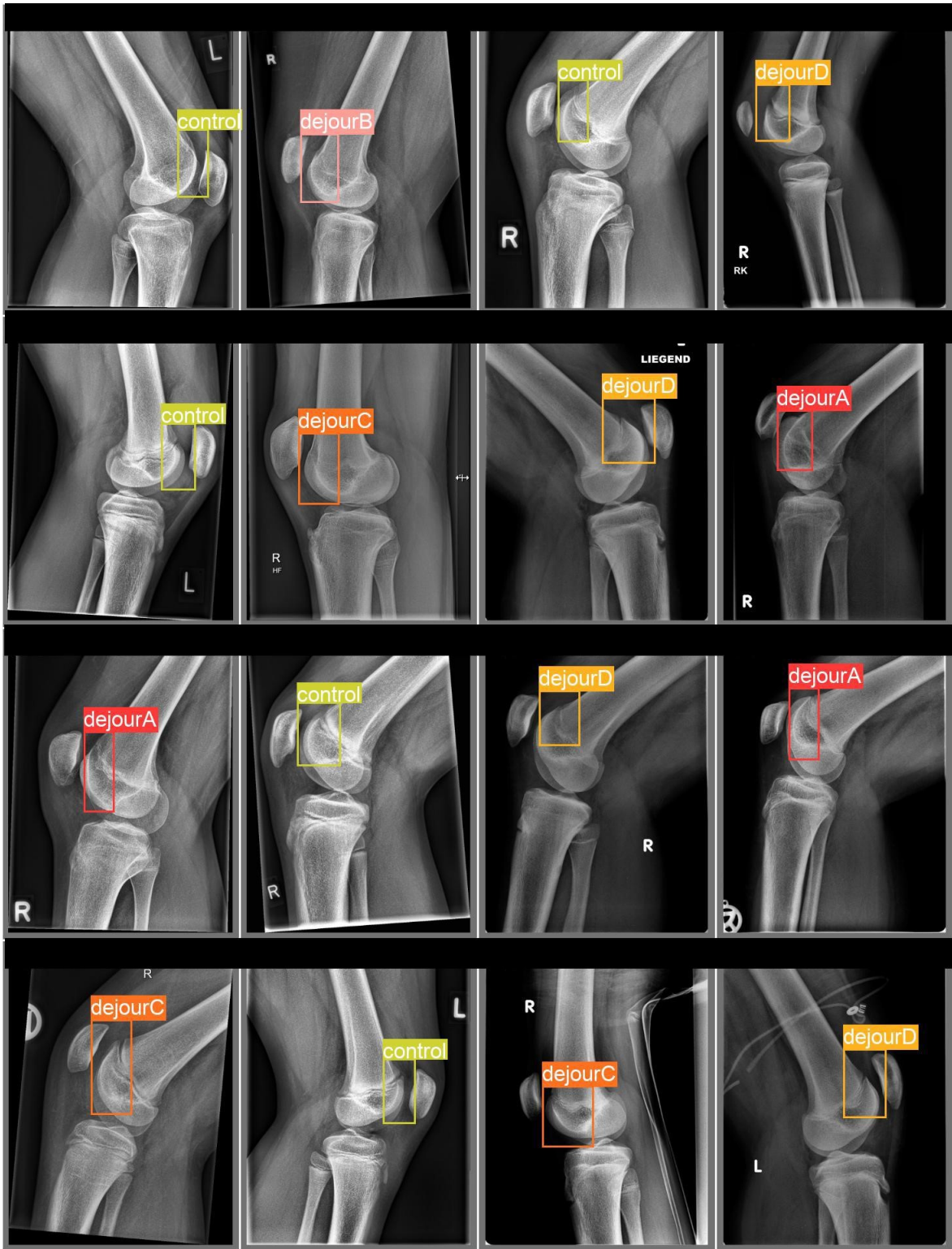


Abbildung 12 Batch2 Labels der Gruppen DejourA-D sowie Control

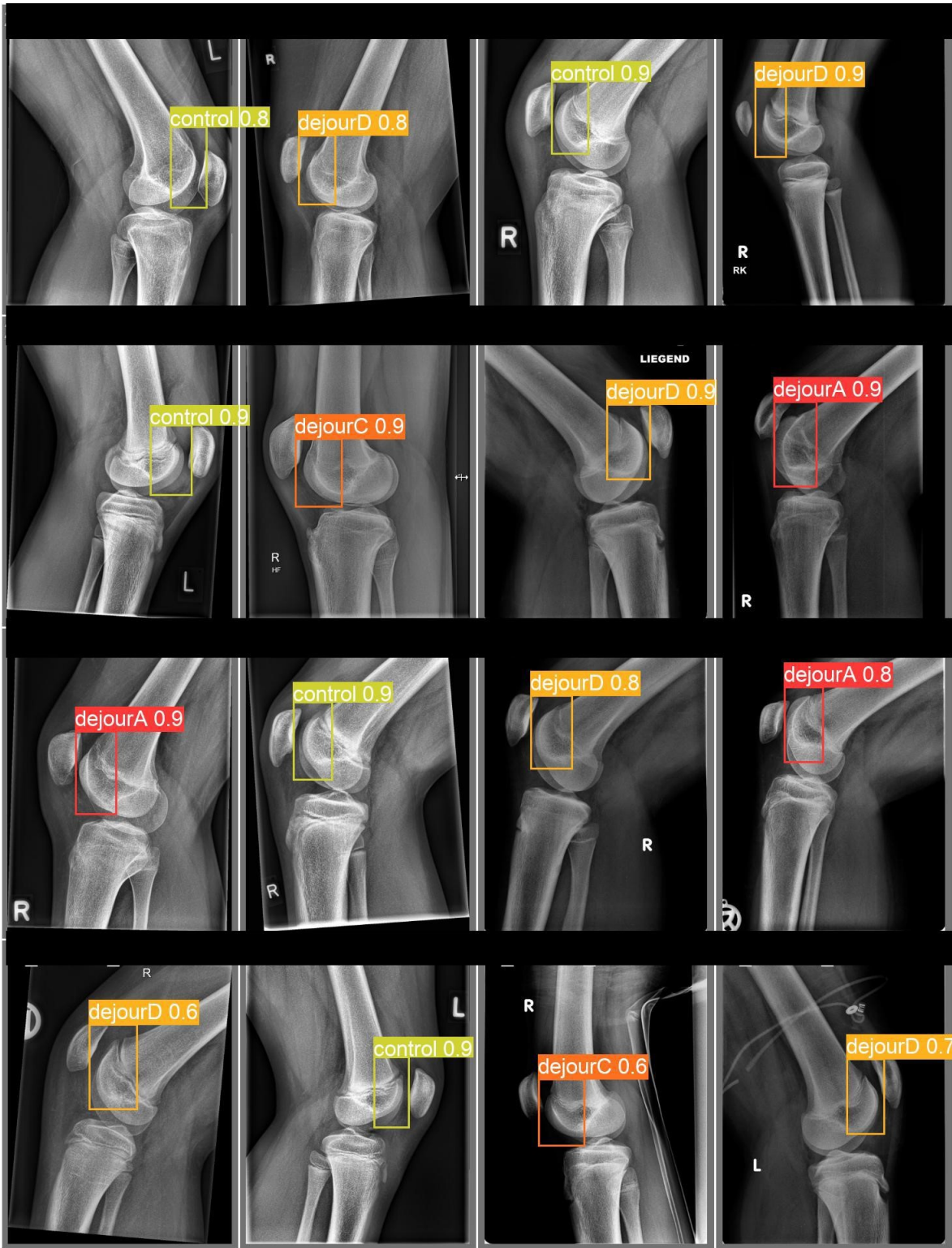


Abbildung 13 Batch2 Labels mit Vorhersagewerten der Gruppen DejourA-D sowie Control

Die oben gezeigten Abbildungen 8, 11 sowie 12 veranschaulichen beispielhaft einen Auszug der vergebenen Labels, die den einzelnen Gruppen Dejour A bis Dejour D sowie

der Gruppe Control zugeordnet wurden. Die Abbildungen 9 sowie 10 und 13 zeigen zusätzlich zu den Labels die zugehörigen Vorhersagewerte, welche vom CNN-Modell YOLOv5 in den Batches 0, 1 und 2 vergeben wurden.

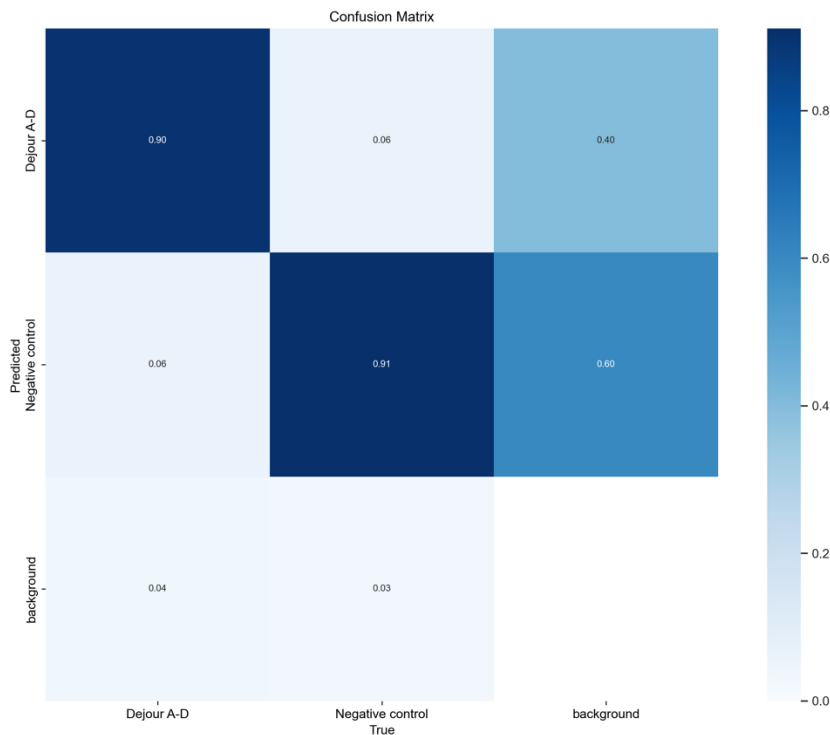


Abbildung 14 Konfusionsmatrix YOLOv5 DejourA-D sowie Gruppe Negative control

Die Abbildung 14 zeigt eine sogenannte Konfusionsmatrix, welche die Gruppen DejourA-D sowie die Gruppe Negative Control in Bezug auf die richtigen Gruppenzugehörigkeiten und die Vorhersagewahrscheinlichkeiten bildlich darstellt. Die Vorhersagewerte sind zusätzlich nochmals in Farbe veranschaulicht, wobei gilt: je dunkler der Farbton, desto höher die Vorhersagewahrscheinlichkeit. So wird deutlich, dass - wenn man die Gruppen DejourA-D zusammenfasst - eine korrekte Vorhersage in 90% der Fälle erreicht werden kann. In der Gruppe der negativen Kontrolle wurden 91% korrekt zugeordnet.

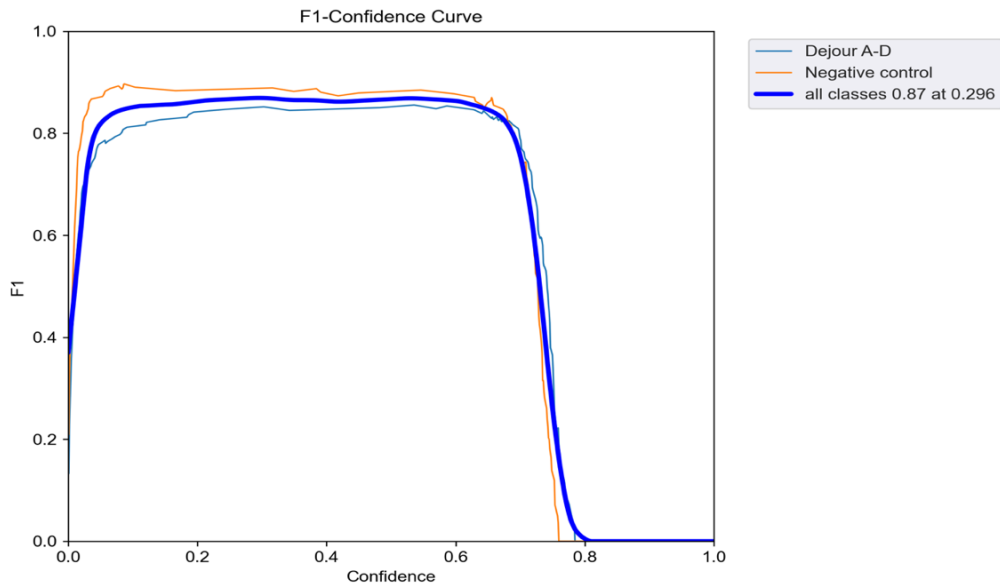


Abbildung 15 F1-Confidence Kurve für zusammengefasste Gruppen DejourA-D

In Abbildung 15 wurden die Klassen zusammengefasst und die Kurve klassenübergreifend dargestellt. Dabei kann über alle Klassen ein F1 Wert von 0,87 bei einer Confidence von 0,296 erreicht werden.

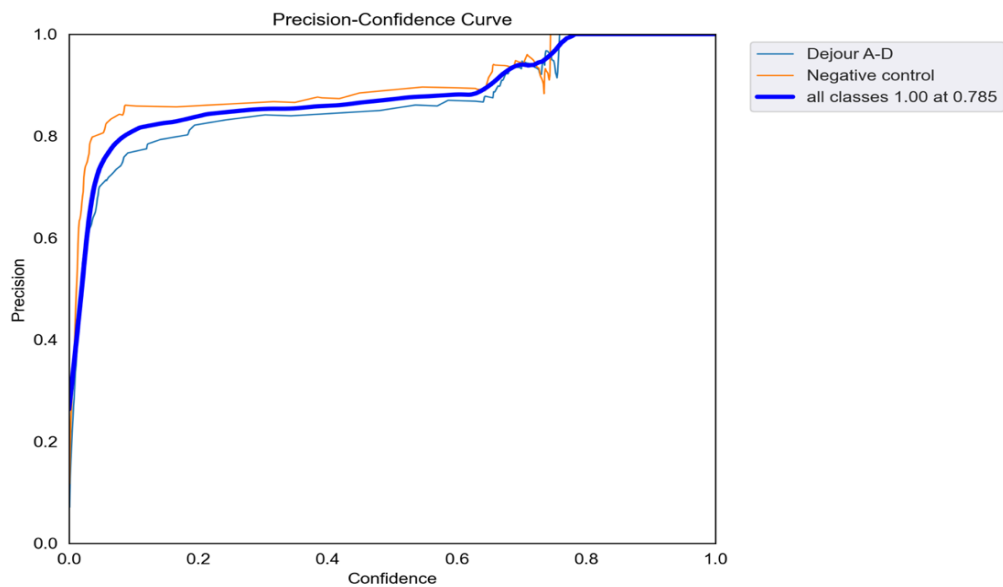


Abbildung 16 Precision Confidence Kurve der Gruppen DejourA-D sowie der Gruppe Negative Control

Führt man die Gruppen DejourA-D zusammen so erhält man oben angeführte Abbildung Nr. 16, welche die Kurven für Precision in Bezug auf Confidence darstellt. Dabei kann für

alle Klassen bei einer Precision von 1,00 ein Wert für Confidence von 0,785 erreicht werden. Dieser Kurvenverlauf ist als dicker Strich dargestellt.

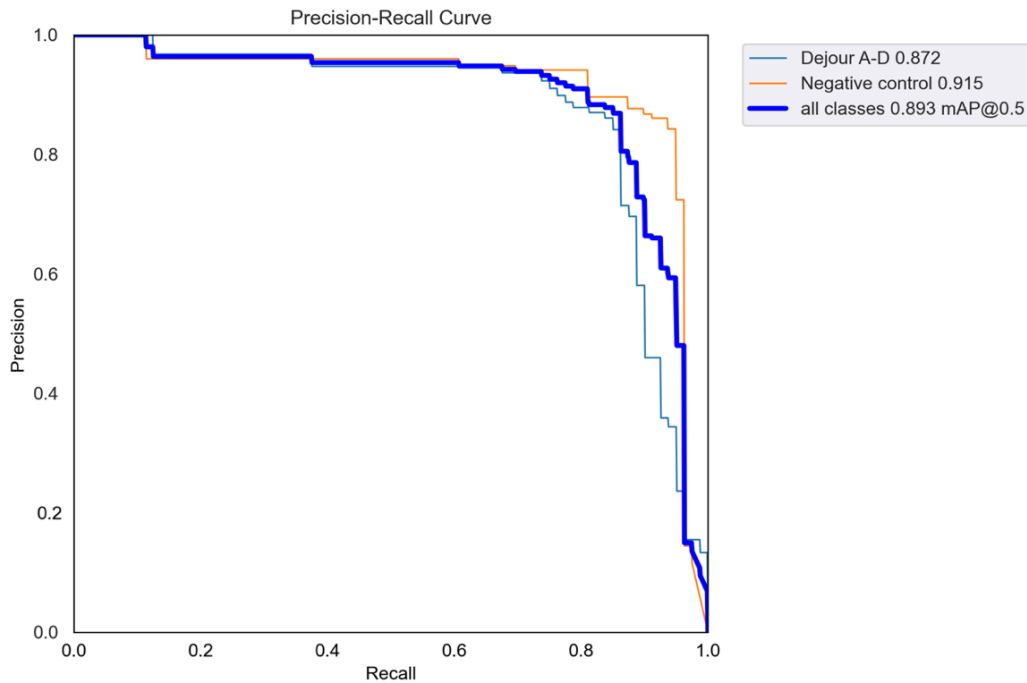


Abbildung 17 Kurvenverläufe für Precision sowie Recall der zusammengefassten Gruppen DejourA-D sowie der Gruppe Negative Control

Die Abbildung 17 stellt wiederum die drei Kurvenverläufe für Precision in Bezug zu Recall dar. Der hervorgehobene blaue Kurvenverlauf stellt dabei die Kurve für alle Klassen dar. Dabei wurde bei einem Wert für Precision 0,893 einen Wert für Recall 0,5 erreicht. Die Gruppe Negative Control ist in der Farbe Orange gezeigt mit Precision 0,915 bei Recall 0,5.

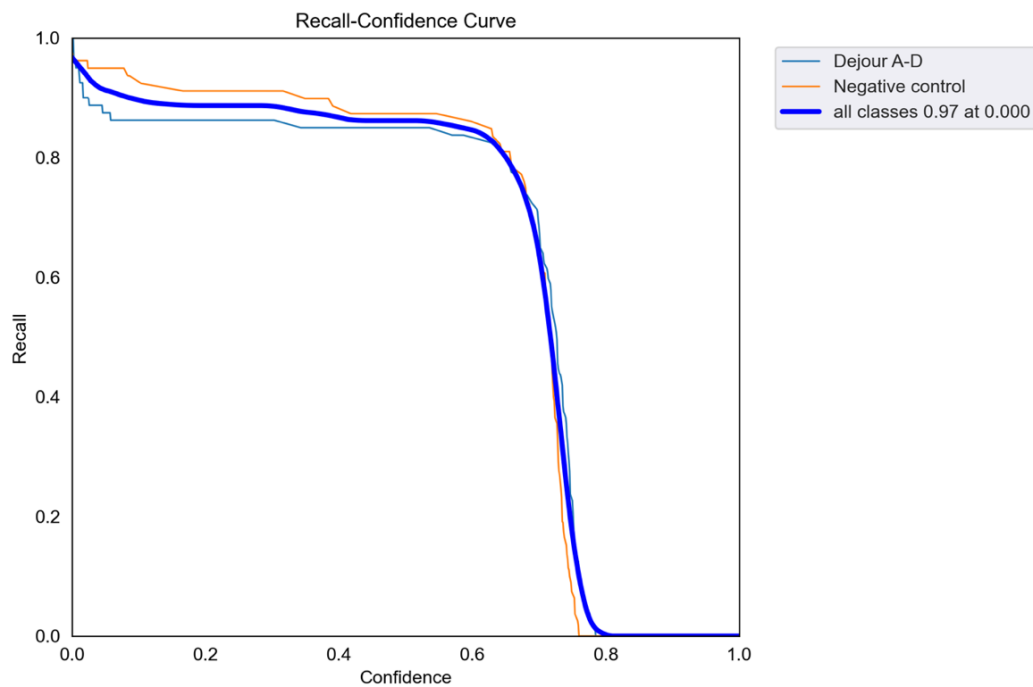


Abbildung 18 Kurvenverläufe für Recall sowie Confidence der zusammengefassten Gruppen DejourA-D sowie der Gruppe Negative Control

Die Abbildung 18 stellt einen Bezug zwischen Recall und Confidence zwischen den zusammengefassten Gruppen DejourA-D sowie der Gruppe Negative Control her. Diesbezüglich kann ein Recall Wert von 0,97 bei Confidence 0,000 erreicht werden. Zugehörige Kurve ist in Abbildung 16 in der Farbe Blau und hervorgehoben dargestellt.

4 Diskussion

Die durchgeführte Studie wurde anhand retrospektiv analysierter Knie Röntgenbilder in seitlicher Projektion durchgeführt. Ziel war die mögliche Erkennung einer Trochleadysplasie nach Dejour mithilfe von 2 verschiedenen CNN-Modellen.

Es zeigte sich, dass CNN-Algorithmen in der Lage waren, eine Trochleadysplasie mit unterschiedlicher Genauigkeit aufzuzeigen. So wurde ein Training mittels selbst erstellter Datenbank und klassifizierten Röntgenbildern nach Dejour A-D und Vergleich mit einer Kontrollgruppe N (keine Pathologie nach Dejour) mit dem CNN-Modell EfficientNet B4 durchgeführt. Die besten Ergebnisse in den Gruppen Dejour A-D wurden in der Gruppe Dejour A erzielt. Insgesamt wurden aber nur 11 der 20 Röntgenbilder richtig klassifiziert also eine Sensitivität von nur 55% erreicht. Die Gruppen Dejour B sowie C und D erreichten noch niedrigere Werte. Die Gruppe Dejour N weist eine Spezifität von 98,75% So können nicht pathologische Röntgenbilder mit sehr hoher Vorhersagewahrscheinlichkeit als nicht pathologisch eingestuft werden, die Unterteilung in die einzelnen pathologischen Dejour Gruppen ist aktuell noch nicht ausgereift und fehleranfällig. Jedoch wurden in allen 4 Gruppen Dejour A-D bei insgesamt 80 präsentierten pathologischen Röntgenbildern nur 14 Bilder als falsch negativ eingeordnet, entsprechend einer Sensitivität von 82,5%. So werden mit dem CNN-Modell annähernd 83% als pathologisch klassifiziert, sofern eine Pathologie vorliegend ist.

Aufgrund einer beschränkten Anzahl an Röntgenaufnahmen der einzelnen Dejour-Grade, war die automatisierte Gruppenzuordnung nur eingeschränkt möglich, wobei die Erkennung der Pathologie im Allgemeinen prinzipiell möglich ist. Durch die Verwendung größerer Datenbanken, könnte es möglich sein, bei der Aufgabe der Gruppenzuordnung bessere Ergebnisse zu erlangen. Da dies derzeit am Standort nicht erzielbar ist, wurde den Gruppenunterscheidungen weniger Priorität beigemessen als der binären Erkennung von normalen und pathologischen Röntgenbildern, entsprechend der Gruppen Dejour A-D (Gruppe YES) sowie nicht pathologischen Röntgenbildern (Gruppe NO). Dem Algorithmus wurden wiederum 80 Bilder pro Gruppe als Test-Datensatz präsentiert. Hierbei konnte eine Spezifität von 98,75% sowie eine Sensitivität von 73,75% erreicht werden. So kann mit sehr hoher Vorhersagerate (98,75%) ein nicht pathologisches Bild

korrekt als nicht pathologisch klassifiziert und mit hoher Vorhersagerate (73,75%) ein pathologisches Röntgenbild als pathologisch eingestuft werden.

Des Weiteren wurde gezeigt, dass mit CNN-Modellen wie beispielsweise EfficientNet-B4 gute Ergebnisse in Unterscheidungen getroffen werden können, ein Einfluss auf die genauen Mechanismen des Lernprozesses selbst beziehungsweise das genaue Unterscheidungsmerkmal nicht vordefiniert werden kann. So kann nicht abschließend geklärt werden, woran das CNN-Modell die Unterscheidung zwischen pathologisch oder nicht pathologisch trifft. So könnten auch andere Merkmale zusätzlich in die Entscheidung miteinbezogen worden sein, wie Knochenalter, Frakturen, Weichteilschatten oder auch Auffälligkeiten an Knochenvorsprüngen wie der Tuberositas tibiae. Um diesem möglichen systematischen Fehler entgegenzuwirken, kann mittels Objekterkennung ein vordefinierter Bereich festgelegt werden, in welchem die Pathologie zu finden ist, um ein genaueres Ergebnis zu erzielen beziehungsweise einen genauen Bereich vorzugeben, anhand welchem das CNN-Modell lernt, und versucht unterschiedliche Merkmale zwischen den pathologischen Röntgenbildern und den nicht pathologischen Röntgenbildern zu finden. Somit kann durch die Objekterkennung der Lernprozess und damit die spätere Klassifikation mittels Festlegen definierter Ausschnitte bereits im Vorhinein besser gesteuert werden. Der Nachteil dieser Variante ist, dass der Aufwand zur Markierung der Daten deutlich zunimmt. In der Regel müssen diese händisch mittels Computer-Software eingezeichnet werden(58). Bei medizinischen Fragestellungen wird dazu in vielen Fällen Fachpersonal mit entsprechender Expertise benötigt. Dieses ist oft zeitlich überlastet oder überhaupt nicht verfügbar, weshalb nicht beliebig Markierungen angefordert oder durchgeführt werden können. Als Alternative hat sich in der Machine Learning Community der sogenannte „Human In The Loop“-Approach (HITL) etabliert (59, 60). Das Prinzip dahinter ist, einen Teil des Datensatzes durch menschliche Expert*innen zu annotieren und anschließend die Ergebnisse mittels Trainings eines neuronalen Netzes auf den verbliebenen Datensatz anzuwenden. Danach muss der nächste Teil des Datensatzes von den Expert*innen lediglich korrigiert werden. Üblicherweise sinkt der Annotierungsaufwand dadurch erheblich. Zum Ende des Datensatzes sollten die vorgeschlagenen Markierungen bereits zufriedenstellende Ergebnisse liefern. Der vorliegende Datensatz bestand aus einer Anzahl an insgesamt 932 Bildern, wobei der

Anteil an normal befundenen Kniegelenken (n=475) keiner speziellen Aufarbeitung bedurfte. Diese Anzahl an Daten kann mit einer geeigneten Lösung zur Bildmarkierung innerhalb kurzer Zeit vollständig annotiert werden. Ein HITL-Zugang wurde deshalb in dieser Arbeit nicht verwendet.

Die Studie hatte folgende Limitationen: geringe Anzahl an Stichproben in den einzelnen Dejour Graden, mögliche menschliche Fehler in der Einteilung der Röntgenbilder in die Dejour-Grade bzw. auch die potenziell fehlerbehaftete Befundung der Kontroll-Röntgenaufnahmen durch Radiolog*innen, die Auswahl der trainierten und getesteten KI-Algorithmen, die Erklärbarkeit der KI-Ergebnisse sowie die Qualität der eingeschlossenen Röntgenaufnahmen.

In diesem Zusammenhang fällt auf, dass die Gruppe Dejour B regelmäßig zu schlechteren Metriken in den Testdurchläufen führte: EfficientNet-B4 30% korrekt, YOLOv5 5% korrekt, schlechte Werte in Bezug auf Confidence, Recall und Precision in Abbildungen 1, 2, 3, 4 und 5. Die Gruppe Dejour B ist die Gruppe, welche in dem gesamten Querschnitt an pathologischen Bildern, am wenigsten stark vertreten war. So waren lediglich 59 Knie Röntgenbilder dieser Gruppe zuzuordnen. Die Vorhersagewahrscheinlichkeit leidet bei derart geringen Datenmengen sinngemäß. Dies lässt sich auch durch die angewandte Datenaugmentation nicht ausgleichen. Augmentation ([61](#)) bezeichnet einen Vorgang vor oder beim Training neuronaler Netze, der das Eingangsbild vervielfacht und die Kopien jeweils verändert. So werden beispielsweise Helligkeit und Kontrast angepasst, das Bild rotiert oder gespiegelt, sowie auch verzerrt, geschärft oder weichgezeichnet. Dadurch kann die Erkennungsrate in vielen Fällen verbessert werden. Um die Vorhersage in dieser Gruppe B sowie auch in anderen Gruppen noch valider zu machen, wäre es aber vorteilhafter, eine größere Menge an Bilddaten mit der entsprechenden Pathologie einzuschließen. So könnten in Zukunft in dieses bereits vortrainierte CNN-Modell weitere Bilder implementiert werden, um bessere Ergebnisse zu erzielen. Denkbar wäre es auch, über eine Multi-Center-Studie mehr Daten aufbereiten und inkludieren zu können.

Die streng seitliche Einstellung beim Knieröntgen ist besonders bei Kindern eine Herausforderung und gelingt nicht in allen Fällen und Situationen gleich gut ([62](#), [63](#)). Die Dejour-Klassifikation bezieht sich aber auf streng seitlich eingestellte Aufnahmen. Die beste Analyse der Bilder gelingt nur, wenn diese Voraussetzung erfüllt ist. Das ist nicht bei

allen Bildern der Fall. Inwieweit diese Abweichungen die CNN-Modelle bei ihrer Klassifikation behinderten, kann abschließend nicht geklärt werden.

Fehler in der Annotierung machen sich als „Rauschen“ in den Daten bemerkbar. Es ist anzunehmen, dass mit einer steigenden Anzahl an Bildern im Datensatz, der Effekt auf das Ergebnis abnimmt. Die manuelle Klassifikation der Röntgenbilder nach Dejour A-D erfordert zudem eine gewisse, Untersucher – bezogene Erfahrung. So ist es nicht auszuschließen, dass Röntgenbilder durch Fehleinschätzungen in eine inkorrekte Gruppe eingeordnet wurden. Die oben angeführte Abweichung der Achse von der streng seitlichen Projektion in den Knie Röntgenbildern kommt zu diesem Umstand erschwerend dazu.

Die Vertiefung in dieses Thema hat bestätigt, dass viele Aspekte und Themen weiterer Bearbeitung und Forschung bedürfen. Diese Pilotstudie hatte die Idee herauszufinden, ob und inwieweit es möglich sein kann, mithilfe von AI-Modellen einen Unterschied zwischen den pathologischen (Dejour A-D klassifizierten) Knie Röntgenbildern (nach Dejour) sowie nicht pathologischen Röntgenbildern herauszuarbeiten. Im Vorhinein war nicht klar, ob es gelingen wird, die KI-Modelle für die spezielle Fragestellung zu trainieren. Als besondere Herausforderung gilt dabei, dass die Entscheidungsfindung des neuronalen Netzwerks nicht genau nachvollzogen werden kann „Black Box“ ([64](#)). Obwohl das Ergebnis korrekt wiedergegeben wird, bleibt unklar, anhand welcher Kriterien die Entscheidung getroffen wurde. Dies ist insbesondere bei Klassifizierungsalgorithmen (z.B. EfficientNet) relevant, da bei Objekterkennungsalgorithmen (z.B. YOLO) die entscheidende Bildregion vorgegeben werden muss. Aus diesem Problem hat sich ein ganzer Zweig der KI-Forschung entwickelt, der an erklärbarer KI („explainable AI“) arbeitet ([65](#), [66](#)). In diesem Kontext kann die Bildklassifikation nicht weiter gesteuert werden, als dass der für das neuronale Netzwerk größte Unterschied zwischen den einzelnen Gruppen gelernt wird. Bei Aufnahmen von Kindern wiegt das doppelt schwer, da auch das Knochenalter einen großen Einfluss auf die Charakteristik der dargestellten Knochenstrukturen hat. Im ungünstigsten Fall würde das neuronale Klassifizierungs-Netzwerk gar nicht den gewünschten Parameter trainieren, sondern andere, für Menschen vielleicht nicht gleich offensichtliche Unterschiede. Werden beispielsweise nur Bilder der rechten Seite in die erste Gruppe und Bilder der linken Seite in die zweite Gruppe

eingespeist, so ist dieser Unterschied höchstwahrscheinlich größer, als geringe Knochenveränderungen an der Trochlea femoris.

Hauptziel für die Zukunft ist, die erstellte Bilddatenbank und deren Markierungen fortlaufend iterativ zu erweitern und zu verbessern, damit nach und nach ein zuverlässiger Algorithmus bereitgestellt werden kann. Hierbei ist es auch interessant, über die Grenzen der Röntgendiagnostik hinaus weiterzudenken.

Die Verfügbarkeit von CT- und MRT-Untersuchungen am Kniegelenk ist besonders groß. Pathologische Befunde oder Messwerte aus diesen Untersuchungen könnten auf die Röntgenbilder übertragen werden, um bisher nicht beachtete Merkmale in den Röntgenbildern zu entdecken. Es ist sehr wahrscheinlich, dass Algorithmen, wie der in dieser Arbeit präsentierte, zukünftig eine große Rolle in der klinischen Routine spielen werden. Es wird zur Normalität gehören, dass Röntgenaufnahmen von einer KI beurteilt werden und Ergebnisse dann zur Verfügung gestellt werden. Dies ist eine große Chance für Patient*innen, da die Röntgendiagnostik nach wie vor sehr von der Expertise der/s jeweiligen Befunder*in abhängt. Oft sind junge Ärzt*innen mit diesen Röntgenuntersuchungen konfrontiert und eine fehlerhafte Beurteilung ist nicht selten ([67-69](#)). Durch die KI könnte eine bessere Standardisierung erfolgen. Außerdem könnte sich die Qualität der Ausbildung dadurch substantiell verbessern.

Zusammenfassend konnte zum ersten Mal erfolgreich Algorithmen zur Unterscheidung der femoropatellaren Trochleadysplasie trainiert werden. Die Ergebnisse sind vielversprechend, aber aufgrund der relativ limitierten Anzahl an inkludierten Röntgenaufnahmen ausbaufähig. Insgesamt konnte aber gezeigt werden, dass die automatisierte Beurteilung der Dejour-Grade mittels KI möglich ist.

Literaturverzeichnis

1. Platzer W, Shiozawa T. Taschenatlas Anatomie, Band 1: Bewegungsapparat 2018.
2. Rixen D SC, Tingart M Kniechirurgie. Elsevier. 2017.
3. Anderhuber F, Pera F, Streicher J. Waldeyer - Anatomie des Menschen: Lehrbuch und Atlas in einem Band: De Gruyter; 2012.
4. Flandry F, Hommel G. Normal anatomy and biomechanics of the knee. Sports Med Arthrosc Rev. 2011;19(2):82-92.
5. Fox AJ, Wanivenhaus F, Burge AJ, Warren RF, Rodeo SA. The human meniscus: a review of anatomy, function, injury, and advances in treatment. Clin Anat. 2015;28(2):269-87.
6. Dejour H, Walch G, Nove-Josserand L, Guier C. Factors of patellar instability: An anatomic radiographic study. Knee Surgery, Sports Traumatology, Arthroscopy. 1994;2(1):19-26.
7. Schmeling A. Trochleadysplasie und ihre Therapie. Knie Journal. 2022;4(2):100-10.
8. Schmeling A, Frings J, Akoto R, Frosch KH. [Patellar dislocation : Causes and treatment]. Unfallchirurg. 2020;123(12):969-83.
9. Dejour H, Walch G, Neyret P, Adeleine P. [Dysplasia of the femoral trochlea]. Rev Chir Orthop Reparatrice Appar Mot. 1990;76(1):45-54.
10. Lippacher S, Reichel H, Nelitz M. Radiological criteria for trochlear dysplasia in children and adolescents. J Pediatr Orthop B. 2011;20(5):341-4.

11. Maas K-J, Warncke ML, Leiderer M, Krause M, Dust T, Frings J, et al. Diagnostic Imaging of Patellofemoral Instability. *Rofo*. 2021;193(09):1019-33.
12. Paech D, Basad E, Kintzelé L. Radiologische Diagnostik und Klassifikationen femuropatellarer Dysplasien. OUP. 2020:132–7.
13. Dejour D, Reynaud P, Lecoultre B. Pain and patellar instability classification. *Med Hyg*. 1998;56:1466-71.
14. Onor G, Yalçın S, Kaar S, Pace I, Ferrua P, Farrow L. The Evaluation of Trochlear Osseous Morphology An Epidemiologic Study. *The Orthopaedic Journal of Sports Medicine*. 2021;9.
15. Kazley JM, Banerjee S. Classifications in Brief: The Dejour Classification of Trochlear Dysplasia. *Clinical orthopaedics and related research*. 2019;477(10):2380-6.
16. Harilainen A, Myllynen P, Antila H, Seitsalo S. The significance of arthroscopy and examination under anaesthesia in the diagnosis of fresh injury haemarthrosis of the knee joint. *Injury*. 1988;19(1):21-4.
17. DeVries CA, Hahn P, Bomar JD, Upasani VV, Pennock AT. Prevalence of trochlear dysplasia in infants evaluated for developmental dysplasia of the hip. *Journal of Children's Orthopaedics*. 2021;15(3):298-303.
18. Pagenstert G, Bachmann M. Klinische Untersuchung bei patellofemorale Problemen. *Der Orthopäde*. 2008;37:890-903.
19. Berrische G, Schmitt H. Die klinische Untersuchung des Kniegelenks. *Deutsche Zeitschrift für Sportmedizin*. 2015;66(03):64-70.

20. Hughston JC. Subluxation of the patella. *J Bone Joint Surg Am.* 1968;50(5):1003-26.
21. Aglietti P, Insall JN, Cerulli G. Patellar pain and incongruence. I: Measurements of incongruence. *Clin Orthop Relat Res.* 1983(176):217-24.
22. Mizuno Y, Kumagai M, Mattessich SM, Elias JJ, Ramrattan N, Cosgarea AJ, et al. Q-angle influences tibiofemoral and patellofemoral kinematics. *J Orthop Res.* 2001;19(5):834-40.
23. Kappel P, Liebensteiner M, Knie-Patellofemoral AGAK. Bildgebende Diagnostik patellofemoraler Störungen. *Arthroskopie.* 2020;33(6):412-9.
24. MRI vs. X-Ray: What You Need to Know [Internet]. 2021. Available from: <https://www.healthline.com/health/mri-vs-xray#takeaway>.
25. Becker C. Konventionelle Röntgendiagnostik der Patella: Einstelltechnik, Tipps und Tricks. *Radiopraxis.* 2016;9(02):91-104.
26. Krüger-Franke M SB, Kugler A, Englert A. Diagnostik und Therapie der akuten Patellaluxation. *Orthopädisch Unfallchirurgische Praxis OUP* 6:284–7.
27. Beaufils P, Thauinat M, Pujol N, Scheffler S, Rossi R, Carmont M. Trochleoplasty in major trochlear dysplasia: current concepts. *Sports Med Arthrosc Rehabil Ther Technol.* 2012;4:7.
28. Sinikumpu J, Nicolaou N. Current concepts in the treatment of first-time patella dislocation in children and adolescents. *J Child Orthop.* 2023;17(1):28-33.
29. Vavken P, Wimmer MD, Camathias C, Quidde J, Valderrabano V, Pagenstert G. Treating Patella Instability in Skeletally Immature Patients. *Arthroscopy: The Journal of Arthroscopic & Related Surgery.* 2013;29(8):1410-22.

30. Longo UG, Ciuffreda M, Locher J, Berton A, Salvatore G, Denaro V. Treatment of Primary Acute Patellar Dislocation: Systematic Review and Quantitative Synthesis of the Literature. *Clin J Sport Med*. 2017;27(6):511-23.
31. Askenberger M, Bengtsson Moström E, Ekström W, Arendt EA, Hellsten A, Mikkelsen C, et al. Operative Repair of Medial Patellofemoral Ligament Injury Versus Knee Brace in Children With an Acute First-Time Traumatic Patellar Dislocation: A Randomized Controlled Trial. *Am J Sports Med*. 2018;46(10):2328-40.
32. Straume-Næsheim TM, Randsborg PH, Mikaelson JR, Årøen A. Medial patellofemoral ligament reconstruction is superior to active rehabilitation in protecting against further patella dislocations. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(10):3428-37.
33. Baier C, Springorum HR, Beckmann J, Grifka J, Matussek J. Therapie der patellaren Instabilität bei Kindern und Jugendlichen. *Der Orthopäde*. 2011;40:868-76.
34. Levy BJ, Tanaka MJ, Fulkerson JP. Current Concepts Regarding Patellofemoral Trochlear Dysplasia. *The American Journal of Sports Medicine*. 2021;49(6):1642-50.
35. Ntagiopoulos PG, Dejour D. Current concepts on trochleoplasty procedures for the surgical treatment of trochlear dysplasia. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2014;22(10):2531-9.
36. Longo UG, Vincenzo C, Mannering N, Ciuffreda M, Salvatore G, Berton A, et al. Trochleoplasty techniques provide good clinical results in patients with trochlear dysplasia. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2018;26(9):2640-58.
37. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9(4):611-29.
38. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. *ArXiv e-prints*. 2015.

39. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks2019.
40. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition2016. 770-8 p.
41. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 14091556. 2014.
42. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284(2):574-82.
43. Yi PH, Kim TK, Wei J, Shin J, Hui FK, Sair HI, et al. Automated semantic labeling of pediatric musculoskeletal radiographs using deep learning. Pediatr Radiol. 2019;49(8):1066-70.
44. Mason DL. pydicom: An open source DICOM library. Available from: <https://github.com/pydicom/pydicom>.
45. Aumasson J-P, Neves S, Wilcox-O’Hearn Z, Winnerlein C, editors. BLAKE2: Simpler, Smaller, Fast as MD5. Applied Cryptography and Network Security; 2013 2013//; Berlin, Heidelberg: Springer Berlin Heidelberg.
46. Morris D. SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions. Federal Inf. Process. Stds. (NIST FIPS), National Institute of Standards and Technology, Gaithersburg, MD; 2015.
47. Nagy E, Janisch M, Hrzić F, Sorantin E, Tschauer S. A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. Scientific Data. 2022;9(1):222.

48. Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B. A survey of modern deep learning based object detection models. *Digital Signal Processing*. 2022;126:103514.
49. 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;35(5):1285-98.
50. Howard Jao. fastai GitHub: GitHub; 2018 [Available from: <https://github.com/fastai/fastai>].
51. Adams SJ, Henderson RDE, Yi X, Babyn P. Artificial Intelligence Solutions for Analysis of X-ray Images. *Can Assoc Radiol J*. 2021;72(1):60-72.
52. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468-73.
53. Tanzi L, Vezzetti E, Moreno R, Aprato A, Audisio A, Massè A. Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach. *Eur J Radiol*. 2020;133:109373.
54. Sekachev B, Manovich N, Zhiltsov M, Zhavoronkov A, Kalinin D, Hoff B, et al. opencv/cvat: v1.1.0. Zenodo; 2020.
55. Jocher; Zhiqiang Wang;Cristi Fati; Jebastin Nadar; Laughing; UnglvKitDe; ACASJBNYKKMTJFLZYCWAVDMG. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation2022.
56. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al., editors. Microsoft COCO: Common Objects in Context. *Computer Vision – ECCV 2014*; 2014 2014//; Cham: Springer International Publishing.

57. Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation 2006. 1015-21 p.
58. Wesp P, Sabel BO, Mittermeier A, Stüber AT, Jeblick K, Schinke P, et al. Automated localization of the medial clavicular epiphyseal cartilages using an object detection network: a step towards deep learning-based forensic age assessment. *Int J Legal Med.* 2023;137(3):733-42.
59. Brostow G. Human in the loop computer vision. *Perception.* 2015;44:360-.
60. Zanzotto FM. Viewpoint: Human-in-the-loop Artificial Intelligence. *J Artif Intell Res.* 2019;64:243-52.
61. Bochkovskiy A, Wang C-Y, Liao H-YM. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:200410934 [cs, eess]. 2020.
62. Thukral BB. Problems and preferences in pediatric imaging. *Indian J Radiol Imaging.* 2015;25(4):359-64.
63. Erondur OF. Challenges and peculiarities of paediatric imaging. *Medical imaging in clinical practice.* 2013;23(23-35):156.
64. Karim MR, Islam T, Shajalal M, Beyan O, Lange C, Cochez M, et al. Explainable AI for Bioinformatics: Methods, Tools and Applications. *Brief Bioinform.* 2023.
65. Sorantin E, Grasser MG, Hemmelmayr A, Tschauer S, Hrzic F, Weiss V, et al. The augmented radiologist: artificial intelligence in the practice of radiology. *Pediatric radiology.* 2022;52(11):2074-86.
66. Gordijn B, Ten Have H. What's wrong with medical black box AI? *Med Health Care Philos.* 2023.

67. Guly HR. Diagnostic errors in an accident and emergency department. *Emerg Med J.* 2001;18(4):263-9.
68. Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department--characteristics of patients and diurnal variation. *BMC emergency medicine.* 2006;6:4.
69. Wei CJ, Tsai WC, Tiu CM, Wu HT, Chiou HJ, Chang CY. Systematic analysis of missed extremity fractures in emergency radiology. *Acta Radiol.* 2006;47(7):710-7.