



OPEN

## Artificial intelligence-based automatic assessment of lower limb torsion on MRI

Justus Schock<sup>1,2,5</sup>, Daniel Truhn<sup>3,5</sup>, Darius Nürnberger<sup>2,3</sup>, Stefan Conrad<sup>4</sup>, Marc Sebastian Huppertz<sup>3</sup>, Sebastian Keil<sup>3</sup>, Christiane Kuhl<sup>3</sup>, Dorit Merhof<sup>2</sup> & Sven Nebelung<sup>3</sup>✉

Abnormal torsion of the lower limbs may adversely affect joint health. This study developed and validated a deep learning-based method for automatic measurement of femoral and tibial torsion on MRI. Axial T2-weighted sequences acquired of the hips, knees, and ankles of 93 patients (mean age,  $13 \pm 5$  years; 52 males) were included and allocated to training ( $n = 60$ ), validation ( $n = 9$ ), and test sets ( $n = 24$ ). A U-net convolutional neural network was trained to segment both femur and tibia, identify osseous anatomic landmarks, define pertinent reference lines, and quantify femoral and tibial torsion. Manual measurements by two radiologists provided the reference standard. Inter-reader comparisons were performed using repeated-measures ANOVA, Pearson's  $r$ , and the intraclass correlation coefficient (ICC). Mean Sørensen–Dice coefficients for segmentation accuracy ranged between 0.89 and 0.93 and erroneous segmentations were scarce. Ranges of torsion as measured by both readers and the algorithm on the same axial image were  $15.8^\circ$ – $18.0^\circ$  (femur) and  $33.9^\circ$ – $35.2^\circ$  (tibia). Correlation coefficients (ranges,  $.968 \leq r \leq .984$  [femur];  $.867 \leq r \leq .904$  [tibia]) and ICCs (ranges,  $.963 \leq \text{ICC} \leq .974$  [femur];  $.867 \leq \text{ICC} \leq .894$  [tibia]) indicated excellent inter-reader agreement. Algorithm-based analysis was faster than manual analysis (7 vs 207 vs 230 s,  $p < .001$ ). In conclusion, fully automatic measurement of torsional alignment is accurate, reliable, and sufficiently fast for clinical workflows.

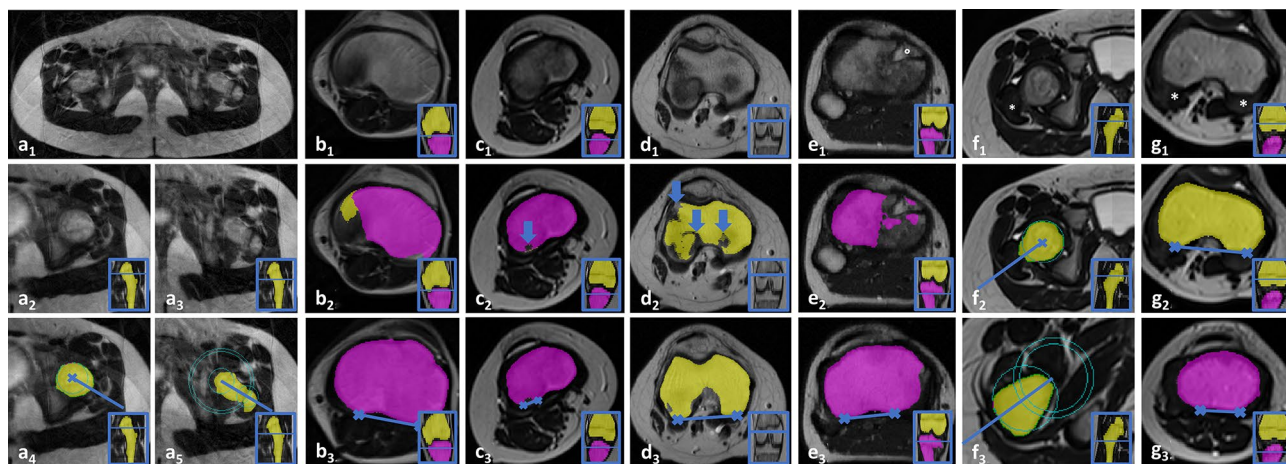
Torsional deformities of the lower limbs are defined as abnormal rotation of the proximal bone axis versus the distal bone axis and may adversely affect joint health directly or indirectly. In addition to hip, knee, and ankle pain, torsional deformities predispose to gait abnormalities such as “in-toeing” or “out-toeing” and are associated with tripping, patellar malalignment and dislocation as well as other osseous deformities that may persist into adulthood<sup>1</sup>. Moreover, torsional deformities have been implicated in the etiopathogenesis of slipped capital femoral epiphysis<sup>2</sup>, femoroacetabular impingement<sup>3</sup>, Legg–Calvé–Perthes disease<sup>4</sup>, developmental dysplasia of the hip<sup>5</sup>, and osteoarthritis<sup>6</sup>.

Exact quantification of torsional deformities is a prerequisite for successful operative treatment and favourable long-term outcomes<sup>7</sup>. Originating in the 1970s, torsional alignment has traditionally been measured by computed tomography (CT)<sup>8</sup>. To reduce radiation exposure to the paediatric patient population, Magnetic Resonance Imaging (MRI) techniques have evolved<sup>9</sup> to be diagnostically equivalent with traditional CT techniques while fitting into tight clinical schedules<sup>10,11</sup>.

For both CT and MRI, accuracy and reliability are challenged by substantial intra- and inter-reader variability that may be as high as  $10.8^\circ$  and  $15.6^\circ$ , respectively<sup>12</sup>, which may be largely attributed to inconsistent level, obliquity, and method of selecting the respective reference lines<sup>8,10</sup>. Yet, even with *a-priori* selected axial or oblique images, the variability persisted<sup>8,13</sup>, highlighting the difficulty of correctly and reliably identifying pertinent axes in 3D objects (such as bone) using 2D images.

In this era of much-sought standardization, there is a clear need for standardized, accurate, and reproducible evaluation of lower limb torsion that may be addressed by deep learning. Deep learning techniques refer to a subtype of machine learning that rely on computational networks to learn from image data by progressively

<sup>1</sup>Department of Diagnostic and Interventional Radiology, University Hospital Düsseldorf, Düsseldorf, Germany. <sup>2</sup>Institute of Imaging and Computer Vision, RWTH University Aachen, Aachen, Germany. <sup>3</sup>Department of Diagnostic and Interventional Radiology, Aachen University Hospital, Pauwels Street 30, 52074 Aachen, Germany. <sup>4</sup>Institute of Informatics, Heinrich Heine University, Düsseldorf, Germany. <sup>5</sup>These authors contributed equally: Justus Schock and Daniel Truhn. ✉email: snebelung@ukaachen.de



**Figure 1.** Representative cases that challenged correct segmentation and determination of reference lines. (a) Motion artefacts during the pelvic MRI scan ( $a_1$ ) caused smearing of the images at the level of the femoral head ( $a_2$ ) and neck ( $a_3$ ), erroneous delineation of the femoral head contour ( $a_4$ ), and incorrect identification of the femoral neck axis through the greater trochanter ( $a_5$ ). (b) Motion artefacts during the MRI scan of the knee ( $b_1$ ) created an artificial protrusion of the medial femur ( $b_2$ ) that did not negatively affect the reference line ( $b_3$ ). (c, d) Prominent and variable epiphyseal plates at the tibia (c) or femur (d) were associated with inaccurate segmentations (block arrows in  $c_2$ ,  $d_2$ ) that caused misplacement of the reference line ( $c_3$ ) or not ( $d_3$ ). (e) Post-surgical changes and fat graft interposition at the anteromedial tibia ( $^\circ$ ) after physiolyis ( $e_1$ ) caused inaccurate segmentation ( $e_2$ ) but did not negatively affect the reference line ( $e_3$ ). (f) At the hip, immature bone characterized by the cartilaginous greater trochanter and femoral neck ( $*$  in  $f_1$ ) is segmented incorrectly ( $f_2$ ) and the femoral neck axis is identified too caudal ( $f_3$ ). (g) Similarly, cartilaginous femoral and tibial condyles ( $*$  in  $g_1$ ) are characterized by incorrect segmentations and reference lines ( $g_2$ ,  $g_3$ ). Inset coronal images framed in blue indicate the height of the axial slices. Patient age and gender: 9 years and male (a, c), 11 years and male (b, e), 12 years and male (d), 5 years and male (g).

	Method	Reader			<i>p</i> -values	Post-hoc test details		
		R1	R2	Alg		Alg vs R1	Alg vs R2	R1 vs R2
Femoral torsion	Lee	16.1 (11.6; 20.6)	18.0 (13.2; 22.9)	15.8 (11.0; 20.6)	<0.001	0.859	<0.001	<0.001
Tibial torsion	Ellipses	33.9 (30.8; 37.0)	34.3 (30.7; 37.9)	35.2 (31.7; 38.6)	0.311	0.212	0.622	0.877

**Table 1.** Mean femoral and tibial torsion as determined by the Lee and ellipses methods. Readers are radiologist 1 (R1), radiologist 2 (R2), and the algorithm (Alg). Means (95% confidence intervals) [ $^\circ$ ]. Statistical analysis was performed by repeated measures ANOVA followed by pair-wise Tukey's post-hoc test. Results of the other methods of determining femoral (according to the Reikeras, Tomczak, and Murphy methods) and tibial torsion (according to the bimalleolar and talus methods) are detailed in Supplementary Table 2. Adjusted *p*-values with significant differences indicated in bold type.

extracting higher-level features without manual engagement. Such techniques have been applied across the disciplines, including medical image processing<sup>14</sup>. In musculoskeletal radiology, the diagnostic possibilities of deep learning are ample and include image reconstruction, image data transformation, tissue segmentation, workflow support, opportunistic screening, compositional analysis, and disease detection<sup>15–21</sup>. To our knowledge, deep learning techniques have not been studied as a diagnostic support system for the assessment of torsional alignment. Against this background, this study's purpose was to develop, train, and validate a deep learning-based diagnostic support system for automatic segmentation and post processing of torsional alignment on MRI. Our hypotheses were that (i) lower limb torsion may be accurately and automatically determined on clinical MRI studies of the hips, knees, and ankles and that (ii) the torsional angles thus obtained are as accurate as those determined manually by radiologists yet at a fraction of the time demand.

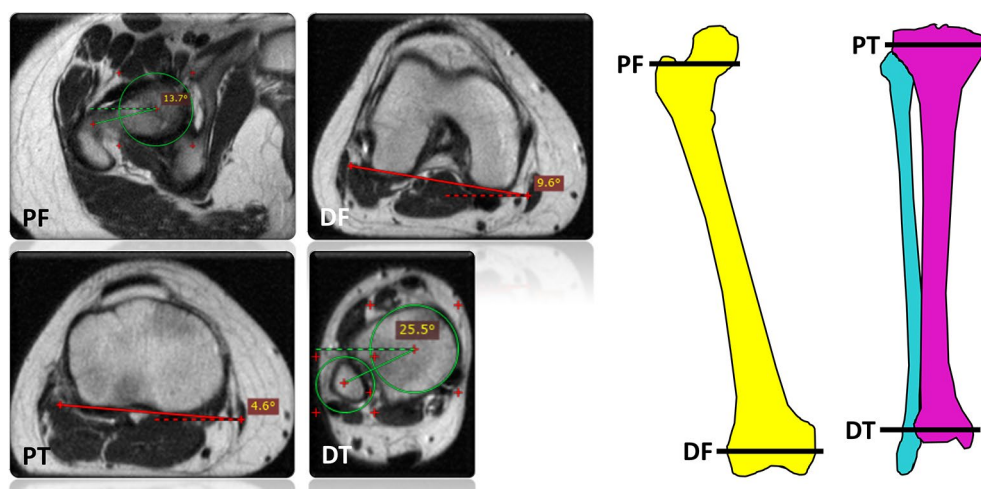
## Results

**Patient characteristics.** Complete MRI studies of 93 patients were included. Patients were children, teenagers, and young adults (mean age, 13.1 years  $\pm$  5.0 [standard deviation]; range 5–34 years; 52 males). After their random allocation to the training ( $n=60$ ), validation ( $n=9$ ), and test sets ( $n=24$ ), demographic details of the sets were largely similar (Supplementary Table 1).

**Segmentation performance.** Similarity of the automatic and manual segmentation outlines was evaluated on the patients of the test set with two lower limbs each. Mean Sørensen-Dice coefficients to quantify simi-

	Method	Comparison	Absolute difference [°]		Pearson's <i>r</i>		ICC	
			R2	Alg	R2	Alg	R2	Alg
Femoral torsion	Lee	R1	2.6 (1.9; 3.3)	3.3 (2.5; 4.0)	0.984 (<0.001)	0.968 (<0.001)	0.974 (0.96; 0.99)	0.963 (0.93; 0.98)
		R2	na	3.7 (3.0; 4.5)	na	0.971 (<0.001)	na	0.966 (0.94; 0.98)
Tibial torsion	Ellipses	R1	5.0 (4.0; 6.0)	4.2 (3.2; 5.1)	0.871 (<0.001)	0.904 (<0.001)	0.865 (0.77; 0.92)	0.867 (0.78; 0.92)
		R2	na	5.1 (4.1; 6.2)	na	0.867 (<0.001)	na	0.894 (0.82; 0.94)

**Table 2.** Metrics of inter-reader difference and agreement for femoral and tibial torsion as determined by the Lee and ellipses methods. Readers were the two radiologists (radiologist 1 [R1], radiologist 2 [R2]) and the algorithm (Alg). In a pair-wise manner, absolute differences [°] and agreement were quantified in terms of Pearson's correlation coefficient *r* and the intraclass-correlation-coefficient (ICC, single scorings [not adjusted]). Means (95% confidence intervals). na – not applicable. Results of the other methods of determining femoral (according to the Reikeras, Tomczak, and Murphy methods) and tibial torsion (according to the bimalleolar and talus methods) are detailed in Supplementary Table 3.



**Figure 2.** Manual reference measurements to determine femoral and tibial torsion at the levels of the hip, knee, and ankle. Anatomic landmarks were used to define the reference lines at the hip in line with the method suggested by Lee (PF proximal femur). The reference lines at the knee were delineated as the distal femoral reference line (DF distal femur) and as the proximal tibial reference line (PT proximal tibia), while the reference line at the ankle was determined using the ellipses method (DT distal tibia). The circles indicate the (superimposed) femoral head and ellipses along the medial malleolus and fibular notch, while dotted lines visualize the horizontal reference lines. Schematics of femur (yellow), tibia (purple), and fibula (light blue) on the right indicate the levels of the axial images. 12-year-old female. Please refer to Supplementary Figure 3 for a visualization of the other manual reference methods to determine femoral torsion (according to the Reikeras, Tomczak, and Murphy methods) and tibial torsion (according to the bimalleolar and talus methods).

ilarity between manually and automatically segmented femoral and tibial outlines were  $0.92 \pm 0.02$ ,  $0.93 \pm 0.03$ ,  $0.93 \pm 0.02$ ,  $0.93 \pm 0.02$ , and  $0.89 \pm 0.02$  for the proximal femur, the distal femur, the proximal tibia, the distal tibia, and the distal fibula, respectively, indicating excellent correspondence in all anatomic regions. Yet, erroneous segmentations occurred if (i) image quality was poor secondary to motion artefacts (Fig. 1a<sub>1-5</sub>, b<sub>1-3</sub>), if (ii) epiphyseal plates were prominent and heterogeneous (Fig. 1c<sub>1-3</sub>, d<sub>1-3</sub>), or if (iii) post-surgical changes were present, e.g., after physiolysis (Fig. 1e<sub>1-3</sub>). Erroneous segmentation outlines were also found in the presence of immature (pre-formed) bone (Fig. 1f<sub>1-3</sub>, g<sub>1-3</sub>). In some instances, incorrect segmentations affected the definition of the reference lines (Fig. 1a<sub>1-5</sub>, c<sub>1-3</sub>, f<sub>1-3</sub>, g<sub>1-3</sub>). At the ankle, all segmentations (48/48 joints) were correct, while at the knee (41/48 joints) and hip (42/48 joints), segmentations were partially incorrect because of excessive motion artefacts (3 joints), variable bone maturation (8 joints), post-surgical status (1 joint), and inhomogeneous bone texture (1 joint).

**Quantification of femoral and tibial torsion.** Mean values of femoral and tibial torsion are indicated in Table 1 and Supplementary Table 2. Mean femoral torsion as determined by both radiologists varied between 16.1° (Lee method, R1) and 32.2° (Murphy method, R1), while the algorithm quantified mean femoral torsion as 15.8°. Mean tibial torsion was equally variable and ranged from 20.9° (talus method, R1) to 34.3° (ellipses

Bone level	Method	Axial image	Description of reference line
Hip/proximal femur	Lee et al. <sup>14</sup>	That visualizes femoral head center, femoral neck, and cephalic junction of greater trochanter (most proximal)	Connection between femoral head center and femoral neck center (single image)
	Reikerås et al. <sup>15</sup>	That visualizes the femoral neck at the level where the anterior and posterior cortices run parallel	Connection between femoral head center and femoral neck axis (superimposed image)
	Tomczak et al. <sup>16</sup>	That visualizes the greater trochanter at the level of the base of the femoral neck	Connection between femoral head center and midpoint of the greater trochanter (superimposed image)
	Murphy et al. <sup>17</sup>	That visualizes the femoral neck's base just cranial to the lesser trochanter	Connection between femoral head center and midpoint of the femoral neck's base
Knee/distal femur	Posterior Condyles <sup>18</sup>	That visualizes the largest diameters of the femoral condyles	Connection between the medial and lateral posterior facets of the femoral condyles
Knee/proximal tibia	Posterior Condyles <sup>18</sup>	That visualizes the tibial condyles just cranial of the fibular head	Connection between the medial and lateral posterior facets of the tibial condyles
Ankle/distal tibia	Ellipses (‡) <sup>8,20</sup>	That visualizes the most distal tibia	Connection of the centers of two ellipses along the medial malleolus and along the fibular notch
	Bimalleolar <sup>8,15</sup>	That visualizes the medial and lateral malleoli and the talar dome	Connection of the centers of the articulating cortices of the medial and lateral malleoli
	Talus <sup>19</sup>	That is just caudal of the tip of the medial malleolus	Line along the anterior surface of the talus

**Table 3.** Description of the methods for manual definition of femoral and distal reference lines as modified from<sup>8,10</sup>. For the reference lines at the knee and ankle, descriptive terminology (instead of the first descriptor's personal name) was used because of as-yet equivocal terminology. Please note that the ellipses method is also referred to as the Ulm method or Waidelech method (‡)<sup>8,36</sup> and that the algorithmic implementation was most similar to the Lee and ellipses methods.

method, R2), with 35.2° determined by the algorithm. Post-hoc testing revealed significant differences between the algorithm and both radiologists for all femoral and tibial methods except for the Lee (Alg vs. R1 [ $p=0.859$ ]) and ellipses methods (Alg vs R1 [ $p=0.212$ ]; Alg vs R2 [ $p=0.622$ ]).

**Inter-reader agreement and differences.** The absolute inter-reader differences between R1 or R2 and the algorithm are detailed in Table 2 and Supplementary Table 3. Because the algorithm was implemented in close resemblance with the Lee and ellipses methods, the absolute inter-reader differences were lowest and the inter-reader correlations ( $0.968 \leq r \leq 0.971$ ) and ICCs ( $0.963 \leq ICC \leq 0.966$ ) were highest for these two methods. With increasing distance to the axial image used for the algorithmic definition of the reference lines, absolute differences increased, and inter-reader ICCs decreased with lowest ICCs found for the Tomczak, Murphy, bimalleolar, and talus methods. Irrespective of the method used for defining the reference lines, inter-reader correlations between R1 or R2 and the algorithm were very strong<sup>22</sup> as indicated by  $0.854 \leq r \leq 0.971$ . Between both radiologists, inter-reader agreement was excellent for femoral torsion ( $ICCs \geq 0.976$ ) and very high for tibial torsion ( $0.871 \leq ICC \leq 0.933$ ), with corresponding mean absolute differences ranging from 2.6° to 3.2° and from 3.3° to 5.0°.

Supplementary Figure 1 visualizes inter-reader differences as a function of their means. No systematic bias in data distribution was observed.

**Time demand.** On average, the algorithm-based analysis of femoral and tibial torsion of one lower limb took 7 s on the specialized workstation, which was significantly faster than the two radiologists (R1:  $207 \pm 17$  s; R2:  $230 \pm 18$  s;  $p < 0.001$ ).

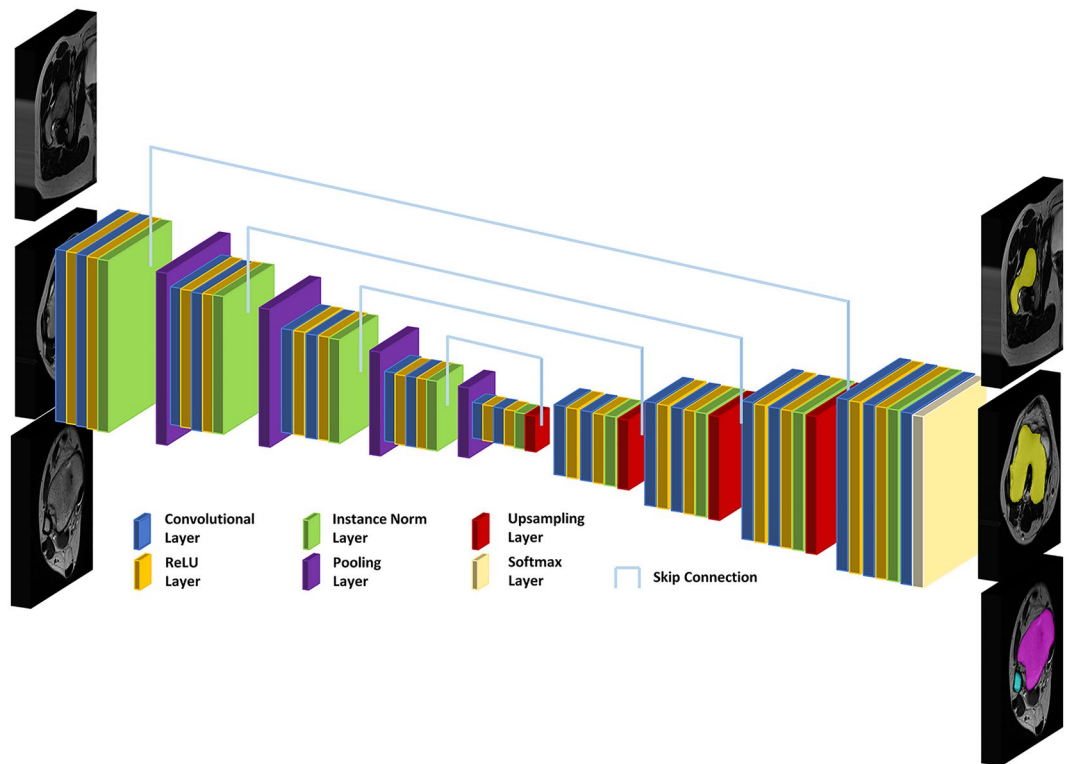
## Discussion

In this study, we implemented an algorithm that automatically analyses lower limb torsion based on MRI. The algorithm performs well in segmenting bone outlines, in identifying relevant anatomy, and in quantifying torsion and thereby renders the automatic measurements of torsional alignment accurate, reliable, stable, and fast for clinical workflows.

Recently, applications of machine learning have exploded across the disciplines, including musculoskeletal radiology, where they are intended to improve clinical workflows<sup>23</sup>. Analysis of torsion seems well suited because image acquisition is standardized<sup>9</sup>, clinical relevance is pronounced<sup>3</sup>, and manual reference methods are variable<sup>8,10</sup>.

With regards to clinical experience and requirements, we intentionally implemented an analysis pipeline of sequential segmentation and post-processing to allow the radiologist to review (and confirm) the outputs. For segmentations, we used a modified, yet conventional U-net convolutional neural network and found excellent segmentation accuracy as indicated by mean Sørensen-Dice coefficients of 0.89–0.93, which is comparable with state-of-the-art techniques<sup>24</sup>. Methodologically, the focus on the Sørensen-Dice coefficient may be justified, yet merely numeric considerations fall short on reflecting the algorithm's performance in challenging anatomies. Segmentations were largely correct yet challenged by poor image quality (due to motion artefacts), heterogeneous epiphyseal plates (due to partial volume effects), post-surgical changes, and immature bone. These conditions were prone to inaccurate segmentations (and, subsequently, inaccurate reference lines) and may be improved



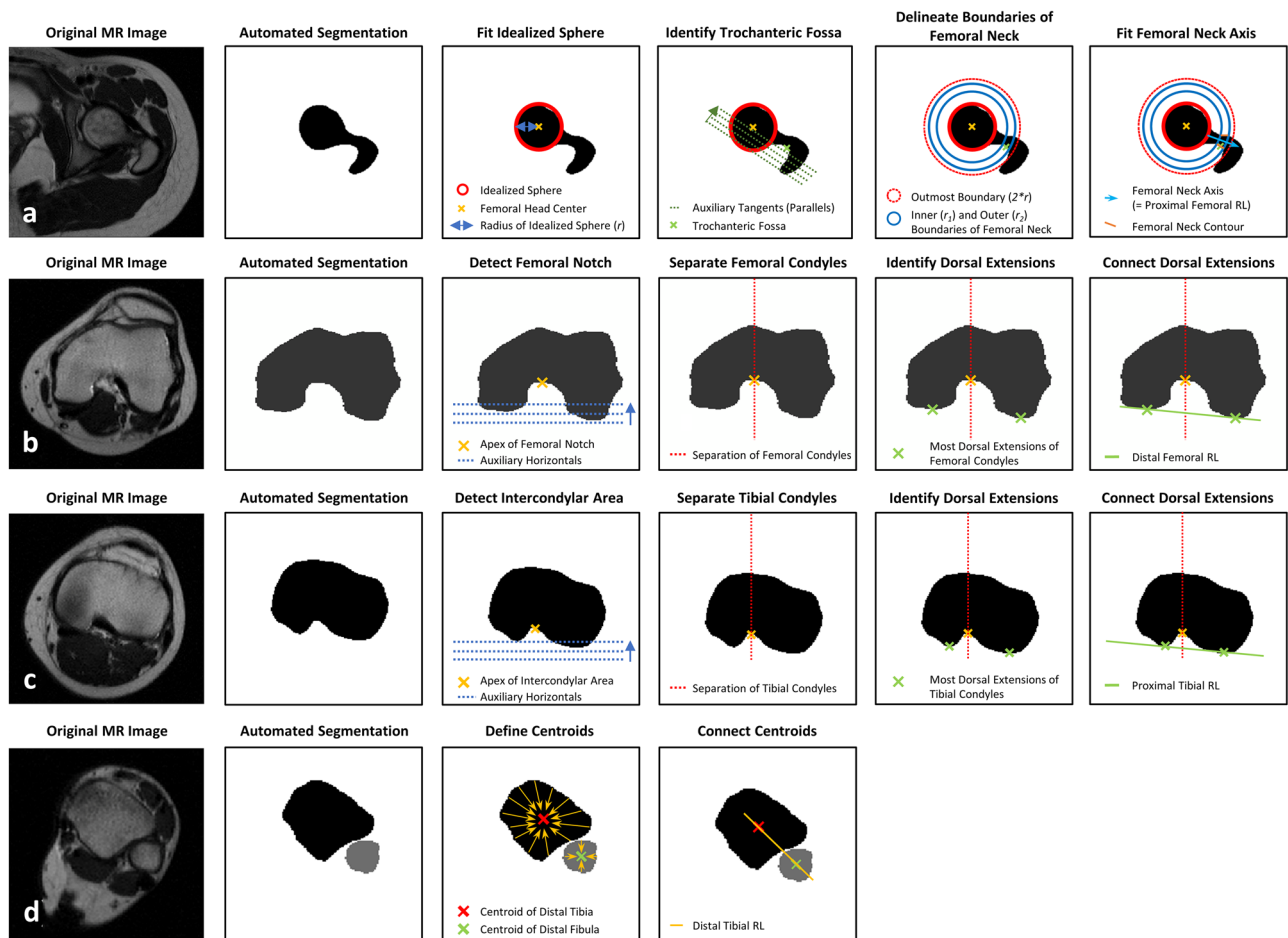


**Figure 3.** Schematic visualization of the U-net convolutional neural network used for automatic segmentations of femur (yellow), tibia (purple), and fibula (light blue). Global information were compressed to a more compact global representation. Local information was preserved by skip connections which passed the more detailed images directly to the decoder path. Up-sampling was performed by bilinear interpolation followed by  $1 \times 1 \times 1$  convolutions. The rectified linear unit (ReLU) was applied in all layers except the last, where a Softmax layer determined each voxels' class probabilities. 3D convolutions aimed to reflect the high dimensionality of the MRI data. Instance normalization was chosen due to small batch size.

by (i) increasing the number and diversity of image datasets, (ii) implementing pre-defined bone-shape models, and (iii) strictly controlling image quality. Including more MRI studies with variable (and dysplastic) limb shapes into the training set will most likely improve the segmentation performance. Yet, not least due to ordering practices of referring orthopedists, overall numbers of torsional MRI studies for adolescents remain low, not just at our institution<sup>10,11</sup>. Including active shape or appearance models may help in generating plausible segmentations. Traditionally, these models have used pre-defined bone shapes that undergo iterative deformation to fit the actual to the ideal bone contours, yet necessitate substantial manual pre-processing<sup>25</sup>. Recent approaches combine shape models with deep-learning techniques<sup>26,27</sup> and may provide viable amendments to our method. Including control mechanisms of image quality may help exclude MRI studies with poor image quality, yet, for the time being, remain the radiologist's responsibility.

Subsequent **algorithmic post-processing** was adapted to the specific anatomy. At the hip, the peripheral femoral neck was automatically identified and used as the proximal reference line. Our method is closest to the method of Lee<sup>28</sup> who also used the most proximal axial image (where the femoral neck is narrowest) to connect the centres of the femoral head and neck. While our method identified the femoral head to locate the peripheral femoral neck, Lee used it as part of the reference line. Also, our method defined the femoral neck axis as the reference line, whereas Lee estimated the femoral neck centre as part of the reference line. Despite these differences, both methods were numerically closely related in terms of mean torsion angles, absolute differences, and inter-reader correlations. Notably, inter-method variability was comparable to inter-reader variability, which underpins the method's reliability. Alternative manual reference methods according to Reikeras, Tomczak, or Murphy<sup>29–31</sup> use more distal axial images for the proximal reference line. Because the distal femoral neck is more posterior (versus the proximal femoral neck), femoral torsion was higher and significantly different from our method. Nonetheless, inter-method correlation in terms of Pearson's  $r$  was excellent, irrespective of the reference method, while ICCs gradually decreased. In our study, Pearson's  $r$  seems the more intuitive measure of association than the ICC, because each value is scaled by its own mean and standard deviation for Pearson's  $r$ , while all data are pooled for the ICC<sup>32</sup>.

Around the knee, the most posterior extensions of the medial and lateral condyles were identified on the axial images with the largest convex bone areas and connected as the reference lines. While referencing to the posterior condyles is largely consented<sup>8–11,28,33</sup> and highly reproducible<sup>8,29</sup>, some discrepancies remain on the

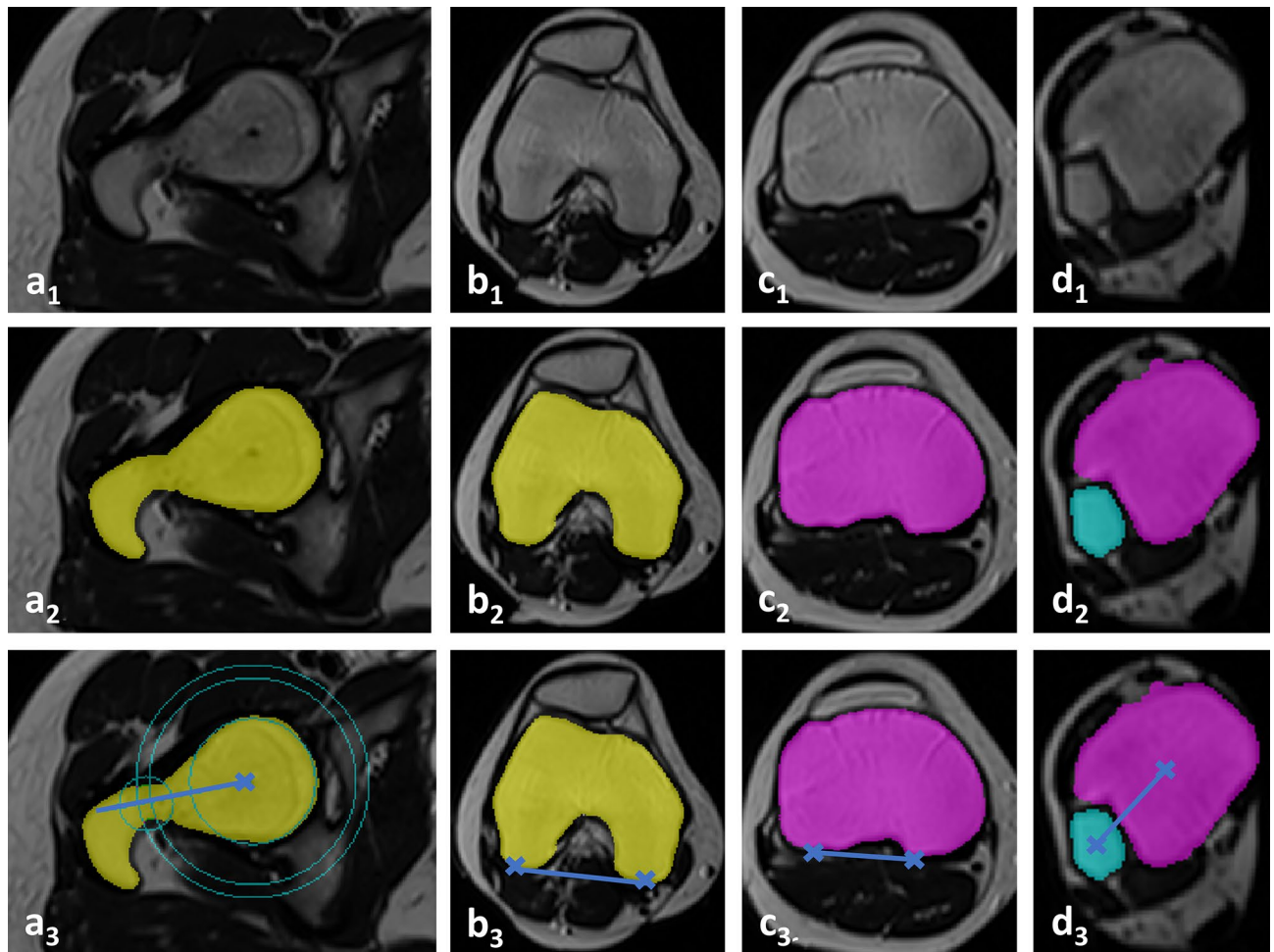


**Figure 4.** Algorithm-based identification of proximal and distal femoral and tibial reference lines to determine lower limb torsion. **(a)** At the hip, the femoral head centre was identified and used to define the femoral neck axis that served as the proximal femoral reference line. **(b, c)** Around the knee, the most posterior extensions of the medial and lateral femoral and tibial condyles were identified and connected as the distal femoral **(b)** and proximal tibial reference lines **(c)**. **(d)** At the ankle, the tibial and fibular centroids were connected as the distal tibial reference line. *RL* reference line. MR images are not to scale.

exact definition of the axial plane<sup>34</sup>. Our method lends itself more to automation than others that involve distance measurements and anatomic considerations<sup>11,34</sup>.

At the ankle, the tibial and fibular centroids were connected as the reference line. Automatically defining the centroids proved reliable and did not result in erroneous segmentations. However, this method may lose accuracy in cases with abnormal or displaced fibulae<sup>35</sup>. Numerically, our method was related closest to the ellipses method that connects the centers of two ellipses placed along the medial malleolus and the fibular notch<sup>8,36</sup>. This finding is plausible as both methods use similar axial images, yet inter-reader agreement was slightly lower than for femoral torsion. Likely, fitting ellipses to non-spherical geometries to define their centers introduces more variability than determining the centroids directly. Alternative manual reference methods such as the bimalleolar and talus methods used more distal axial images<sup>8,37</sup> and were characterized by lower tibial torsion. In line with earlier studies<sup>8,34,35</sup>, these findings reflect the different axes of the talocrural joint<sup>38</sup> and highlight the value of methodologic consistency. Most likely, the lower ICCs (and very high Pearson's  $r$ ) for the distal methods are indicative of differences in reference images, lines, and anatomy.

This study has **limitations**. First, patient numbers were limited which is mainly due to conservative ordering practices that is still heavily reliant on CT and reduces the availability of MRI studies. Second, we only assessed T2-weighted sequences that had been acquired using a specific scanner-coil-configuration at a single site, while other institutions may use different settings so that variability across vendors, platforms, and sequences remains unaccounted for. Consequently, the algorithm's generalizability (and multi-institutional validation) remain to be addressed. Third, even though consecutive patients from the clinical routine were included (and no patients were excluded for medical reasons), our results relate to younger patients with non-traumatic torsional deformities only, while post-traumatic conditions were not assessed. Fourth, alternative deep learning-based methods that derive predictions of torsional alignment directly from the MR images (without necessitating additional bone segmentation, landmark identification, and reference line definition as intermediate steps) were not evaluated and may improve predictions.



**Figure 5.** Visualization of the algorithm-based identification of the reference lines. Indicated are the original MR images ( $a_1$ – $d_1$ ), their segmentation outlines on selected axial images ( $a_2$ – $d_2$ ), and the corresponding reference lines ( $a_3$ – $d_3$ ) as determined by the algorithm. Proximal femur (**a**), distal femur (**b**), proximal tibia (**c**), and distal tibia (**d**). Reference lines and points are blue. For algorithmic details please refer to Fig. 4. Femur (yellow), tibia (purple), and fibula (light blue). 19-year-old female.

In **conclusion**, femoral and tibial torsion can be automatically assessed on clinical MR images of the hip, knee, and ankle joints based on U-net convolutional neural networks for segmentation and algorithm-based analyses of anatomic landmarks and reference lines. The developed methodology is sufficiently accurate, reliable, and fast to enhance clinical workflows.

## Materials and methods

**Study design.** This is a retrospective single-center comparative imaging study that was conducted in accordance with local data protection regulations. The local ethical committee (Ethical Committee of the Medical Faculty, RWTH [Rheinisch-Westfälische Technische Hochschule] Aachen University, reference no. 028/19) approved the study and waived the requirement to obtain individual informed consent. All pediatric and adult patients who had undergone torsional MRI, consisting of stacks through the pelvis, both knees, and both ankles, between July 2013 and August 2020 were identified by a database query. Of the 107 patients thus identified, 14 were excluded because at least one joint was missing ( $n=8$ ) or motion artefacts were excessive ( $n=6$ ). The flowchart is visualized in Supplementary Figure 2. Otherwise, no further exclusion criteria were defined and patients with skeletal dysplasia, other skeletal abnormalities, and orthopedic hardware were included to reflect clinical reality. Eventually, datasets of 93 patients were included and randomly allocated to strictly separated training ( $n=60$ , 65%), validation ( $n=9$ , 10%), and test sets ( $n=24$ , 26%). On the training set, the algorithms were developed and trained to perform image segmentation and post processing, while on the validation set, the optimal hyperparameters that determine the network's structure and training were validated. On the test set, the optimized algorithm's performance was evaluated.

**Image acquisition.** MR images were acquired on a clinical 3.0 T MRI scanner (Achieva, Philips). Neutral position of the supine patient was maintained by extending the knee joints with the patellae facing as anterior as possible while position was maintained by sandbags and positioning aids. The inbuilt body coil was used

for streamlined image acquisition from pelvis to feet while eliminating the need for repositioning. Following a triplane localizer sequence, axial T2-weighted non-fat-saturated 2D turbospin-echo sequences were acquired in three stacks over the pelvis (both hips), knees, and ankles using the protocol detailed in Supplementary Table 4. Centered on the joint, 29 slices were acquired to fully cover each joint.

**Manual reference measurements.** Two clinical radiologists (M.H. and S.N., with 1 and 8 years of experience in musculoskeletal imaging) performed the manual reference measurements on the anonymized test set using the RadiAnt DICOM Viewer (version 2020.1.1, Medixant) and its standard image analysis toolbox. Established methods of reference were chosen as visualized in Fig. 2 and Supplementary Figure 3. At the hip, the reference lines were defined from proximal to distal according to Lee<sup>28</sup>, Reikerås<sup>30</sup>, Tomczak<sup>31</sup>, and Murphy<sup>29</sup>, while at the knee, the posterior femoral and tibial condyles were connected<sup>33</sup>. At the ankle, reference lines were identified using the ellipses, bimalleolar, and talus methods<sup>8,36,37</sup>. The individual methods to define the reference lines are described in Table 3.

Projected angles between the proximal and distal reference lines were determined to calculate torsion. Internal (or external) rotation of the distal relative to the proximal bone is reflective of antetorsion (or retrotorsion)<sup>9</sup>. Also, radiologists' time demand for bilateral measurements of femoral and tibial torsion based on the Lee and ellipses methods was measured on eight random MRI studies.

**Manual segmentations.** The bone contours of the femur, tibia, and fibula were manually delineated on all slices as 'ground truth' segmentations. DN (student of the imaging sciences) labeled the bone contours using ITK-SNAP software (v3.8, Cognitica; <https://www.itksnap.org>) and its function of semiautomatic segmentation<sup>39</sup> after being trained on five MRI studies by an experienced musculoskeletal radiologist (SN) who also reviewed all segmentation outlines for accuracy and consistency.

**Algorithm-based segmentations.** For automatic segmentations, we used a U-net convolutional neural network as suggested previously<sup>40</sup>. At a depth of 6, the network was designed with 16 filters per convolution that doubled with each pooling layer (Fig. 3). Minor modifications to the original network architecture<sup>40</sup> included (i) instance normalization instead of batch normalization (due to small batch sizes), (ii) trilinear interpolation for up-sampling followed by  $1 \times 1 \times 1$  convolution instead of transposed convolutions (to reduce parameter numbers), and (iii) 3D convolutions instead of 2D convolutions (to reflect the higher dimensionality of the MRI data). Data pre-processing was implemented based on a self-adapting framework that yields excellent segmentation results<sup>24</sup> and included resampling to the median voxel spacing of  $0.6 \times 0.6 \times 6.5 \text{ mm}^3$ , cropping, and padding. The voxels' signal intensity distribution was normalized to a mean of 0 and a standard deviation of 1 across the whole image.

During training, modelled predictions of segmentations were compared to the ground truth by employing a combination of the soft Dice loss on all foreground classes, i.e., femur, tibia, and fibula, and the categorical cross-entropy loss on all classes<sup>24</sup>. Calculated loss values were used to optimize the network parameters using stochastic gradient descent optimization with *adaptive moment estimation* (Adam)<sup>41</sup> and a learning rate scheduling system that avoids local minima by decreasing the learning rate by two on training plateaus through monitoring of the moving training loss average. To avoid overfitting, data were augmented by including a random choice of motion, ghosting, spiking, and bias-field artifacts<sup>42</sup>. Training was performed on a *state-of-the-art* graphic-processing-unit (GeForce RTX 3090, NVIDIA) with a batch size of 1 and an initial learning rate of 0.001.

**Algorithm-based analysis of torsion.** After segmentations, femoral and tibial torsion were determined based on the proximal and distal reference lines and used to quantify each bone's torsion. The algorithmic procedures are schematically visualized in Fig. 4 and described in detail in the Supplementary Text 1. In a representative patient, Fig. 5 visualizes the original MR images, the automatic segmentation outlines, and the corresponding reference lines.

**Algorithm availability and time demand.** The algorithm is publicly available on GitHub (<https://github.com>). Time demand for analysis of torsion in one lower limb (consisting of three MRI stacks) was determined on a specialized workstation with a dedicated graphic-processing-unit (Intel-Core i7-9700 K@3.60 GHz, GeForce RTX 2080 Ti, NVIDIA) and evaluated against the manual reference measurements using one-way analysis of variance (ANOVA).

**Statistical analysis.** Statistical analysis was performed by J.S. and S.N. using the Python libraries *statsmodels* and *NumPy*, GraphPad Prism software (version 9.1.1), and R software (version 4.0.2). The Sørensen-Dice coefficient quantified correspondence of manual and automatic segmentation outlines of femur, tibia, and fibula. The D'Agostino-Pearson omnibus normality test was used to confirm underlying normality in torsional values. Torsional angles are reported for each reader, i.e., radiologist 1 (R1), radiologist 2 (R2), and the algorithm (Alg), and for the different methods that were used to select the proximal femoral and distal tibial reference lines. Absolute inter-reader differences were determined using absolute values. Inter-reader correlations and agreement were determined using Pearson's correlation coefficient  $r$  and the intraclass-correlation-coefficient (ICC, single scorings, not adjusted). Repeated measures ANOVA followed by Tukey's *post-hoc* tests were used for pairwise inter-reader comparisons. Multiplicity-adjusted  $p$ -values are reported throughout to account for multiple comparisons against the family-wise alpha error threshold of  $p \leq 0.05$ .



## Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. Any additional datasets generated and analyzed in this study are available from the corresponding author on reasonable request.

## Code availability

The software routines developed, validated, and implemented are publicly available on GitHub (<https://github.com>). Any other data for research purposes are available from the corresponding author upon reasonable request.

Received: 6 September 2021; Accepted: 11 November 2021

Published online: 01 December 2021

## References

1. Gruskay, J. A., Fragomen, A. T. & Rozbruch, S. R. Idiopathic rotational abnormalities of the lower extremities in children and adults. *JBJS Rev.* **7**, e3. <https://doi.org/10.2106/JBJS.RVW.18.00016> (2019).
2. Gelberman, R. H. *et al.* The association of femoral retroversion with slipped capital femoral epiphysis. *J. Bone Joint Surg. Am.* **68**, 1000–1007 (1986).
3. Eijnisman, L. *et al.* Relationship between femoral anteversion and findings in hips with femoroacetabular impingement. *Orthopedics* **36**, e293–300. <https://doi.org/10.3928/01477447-20130222-17> (2013).
4. Upadhyay, S. S., Burwell, R. G. & Moulton, A. Femoral anteversion in Perthes' disease with observations on irritable hips. Application of a new method using ultrasound. *Clin. Orthop. Relat. Res.* **209**, 70–76 (1986).
5. Argenson, J. N., Flecher, X., Parratte, S. & Aubaniac, J. M. Anatomy of the dysplastic hip and consequences for total hip arthroplasty. *Clin. Orthop. Relat. Res.* **465**, 40–45. <https://doi.org/10.1097/BLO.0b013e3181576052> (2007).
6. Tonnis, D. & Heinecke, A. Acetabular and femoral anteversion: relationship with osteoarthritis of the hip. *J. Bone Joint Surg. Am.* **81**, 1747–1770. <https://doi.org/10.2106/0004623-199912000-00014> (1999).
7. Nelitz, M. Femoral derotational osteotomies. *Curr. Rev. Musculoskel. Med.* **11**, 272–279 (2018).
8. Liodakis, E. *et al.* Reliability of the assessment of lower limb torsion using computed tomography: analysis of five different techniques. *Skel. Radiol.* **41**, 305–311 (2012).
9. Koenig, J. K., Pring, M. E. & Dwek, J. R. MR evaluation of femoral neck version and tibial torsion. *Pediatr. Radiol.* **42**, 113–115 (2012).
10. Schmaranzer, F. *et al.* The effect of modality and landmark selection on MRI and CT femoral torsion angles. *Radiology* **296**, 381–390 (2020).
11. Hesham, K. *et al.* Measurement of femoral version by MRI is as reliable and reproducible as CT in children and adolescents with hip disorders. *J. Pediatr. Orthop.* **37**, 557 (2017).
12. Jaarsma, R. *et al.* Computed tomography determined femoral torsion is not accurate. *Arch. Orthop. Trauma Surg.* **124**, 552–554 (2004).
13. Fuller, C. B. *et al.* Femoral version: comparison among advanced imaging methods. *J. Orthop. Res.* **36**, 1536–1542 (2018).
14. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
15. Kalra, A., Chakraborty, A., Fine, B. & Reicher, J. Machine learning for automation of radiology protocols for quality and efficiency improvement. *J. Am. Coll. Radiol.* **17**, 1149–1158. <https://doi.org/10.1016/j.jacr.2020.03.012> (2020).
16. Medina, G., Buckless, C. G., Thomasson, E., Oh, L. S. & Torriani, M. Deep learning method for segmentation of rotator cuff muscles on MR images. *Skeletal Radiol.* **50**, 683–692. <https://doi.org/10.1007/s00256-020-03599-2> (2021).
17. Liu, F. *et al.* Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* **289**, 160–169 (2018).
18. Fritz, J., Guggenberger, R. & Del Grande, F. Rapid musculoskeletal MRI in 2021: clinical application of advanced accelerated techniques. *AJR Am. J. Roentgenol.* **216**, 718–733. <https://doi.org/10.2214/AJR.20.22902> (2021).
19. Bien, N. *et al.* Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med.* **15**, e1002699. <https://doi.org/10.1371/journal.pmed.1002699> (2018).
20. Yang, S. *et al.* Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. *Clin. Radiol.* **75**, 713. <https://doi.org/10.1016/j.crad.2020.05.021> (2020).
21. Fritz, J., Kijowski, R. & Recht, M. P. Artificial intelligence in musculoskeletal imaging: a perspective on value propositions, clinical use, and obstacles. *Skeletal Radiol.* <https://doi.org/10.1007/s00256-021-03802-y> (2021).
22. Evans, J. D. *Straightforward statistics for the behavioral sciences.* (Thomson Brooks/Cole Publishing Co, 1996).
23. Chea, P. & Mandell, J. C. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal Radiol.* **49**, 183–197 (2020).
24. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
25. Baldwin, M. A., Langenderfer, J. E., Rullkoetter, P. J. & Laz, P. J. Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach. *Comput. Methods Prog. Biomed.* **97**, 232–240 (2010).
26. Ambellan, F., Tack, A., Ehlke, M. & Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Med. Image Anal.* **52**, 109–118. <https://doi.org/10.1016/j.media.2018.11.009> (2019).
27. Schock, J. *et al.* in *International Workshop on Shape in Medical Imaging*. 85–94 (Springer).
28. Lee, Y. S., Oh, S. H., Seon, J. K., Song, E. K. & Yoon, T. R. 3D femoral neck anteversion measurements based on the posterior femoral plane in ORTHODOC® system. *Med. Biol. Eng. Comput.* **44**, 895 (2006).
29. Murphy, S. B., Simon, S. R., Kijewski, P. K., Wilkinson, R. H. & Griscom, N. T. Femoral anteversion. *J. Bone Joint Surg. Am.* **69**, 1169–1176 (1987).
30. Reikerås, O., Bjerkeim, I. & Kolbenstvedt, A. Anteversion of the acetabulum and femoral neck in normals and in patients with osteoarthritis of the hip. *Acta Orthop. Scand.* **54**, 18–23 (1983).
31. Tomczak, R. J. *et al.* MR imaging measurement of the femoral antetorsional angle as a new technique: comparison with CT in children and adults. *AJR Am. J. Roentgenol.* **168**, 791–794 (1997).
32. Armstrong, R. A. Should Pearson's correlation coefficient be avoided?. *Ophthalm. Physiol. Opt.* **39**, 316–327 (2019).
33. Waldt, S., Woertler, K. & Telger, T. C. *Measurements and classifications in musculoskeletal radiology* (Thieme Stuttgart, 2014).
34. Hoch, A. *et al.* Tibial torsion analysis in computed tomography: development and validation of a real 3D measurement technique. *Insights Imag.* **12**, 1–7 (2021).
35. Madadi, F. *et al.* A new method for tibial torsion measurement by computerized tomography. *J. Orthop.* **13**, 43–47 (2016).
36. Waidelich, H.-A., Strecker, W. & Schneider, E. in *Posttraumatische Beindeformitäten* 22–29 (Springer, 1997).

37. Basaran, S. H. *et al.* The measurement of tibial torsion by magnetic resonance imaging in children: the comparison of three different methods. *Eur. J. Orthop. Surg. Traumatol.* **25**, 1327–1332 (2015).
38. Claassen, L. *et al.* The geometrical axis of the talocrural joint—suggestions for a new measurement of the talocrural joint axis. *Foot Ankle Surg.* **25**, 371–377 (2019).
39. Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
40. Ronneberger, O., Fischer, P. & Brox, T. in *International Conference on Medical image computing and computer-assisted intervention*. 234–241 (Springer).
41. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. (2014).
42. Pérez-García, F., Sparks, R. & Ourselin, S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. (2020).

## Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (Grant Number: NE 2136/3-1) and the START Program of the Faculty of Medicine, RWTH Aachen University, Germany. The acknowledged sponsors had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the manuscript for publication.

## Author contributions

The contributions of the authors were as follows: Guarantors of integrity of entire study, J.S., D.T., S.N.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, J.S., D.T., S.N.; experimental studies, J.S., D.T., D.N., M.H., S.N.; manuscript editing J.S., D.T., S.C., S.K., C.K., D.M., S.N.

## Funding

Open Access funding enabled and organized by Projekt DEAL. Deutsche Forschungsgemeinschaft, NE 2136/3-1, Medizinische Fakultät, RWTH Aachen University.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02708-y>.

**Correspondence** and requests for materials should be addressed to S.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021