



Deep-Learning System for Automatic Measurement of the Femorotibial Rotational Angle on Lower-Extremity Computed Tomography

Sheen-Woo Lee¹ · Gi Pyo Lee² · Jeun Yoon¹ · Young Jae Kim³ · Kwang Gi Kim^{2,4}

Received: 16 April 2025 / Revised: 19 July 2025 / Accepted: 1 August 2025
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2025

Abstract

To develop and validate a deep-learning-based algorithm for automatic identification of anatomical landmarks and calculating femoral and tibial version angles (FTT angles) on lower-extremity CT scans. In this IRB-approved, retrospective study, lower-extremity CT scans from 270 adult patients (median age, 69 years; female to male ratio, 235:35) were analyzed. CT data were preprocessed using contrast-limited adaptive histogram equalization and RGB superposition to enhance tissue boundary distinction. The Attention U-Net model was trained using the gold standard of manual labeling and landmark drawing, enabling it to segment bones, detect landmarks, create lines, and automatically measure the femoral version and tibial torsion angles. The model's performance was validated against manual segmentations by a musculoskeletal radiologist using a test dataset. The segmentation model demonstrated $92.16\% \pm 0.02$ sensitivity, $99.96\% \pm <0.01$ specificity, and 2.14 ± 2.39 HD95, with a Dice similarity coefficient (DSC) of $93.12\% \pm 0.01$. Automatic measurements of femoral and tibial torsion angles showed good correlation with radiologists' measurements, with correlation coefficients of 0.64 for femoral and 0.54 for tibial angles ($p < 0.05$). Automated segmentation significantly reduced the measurement time per leg compared to manual methods (57.5 ± 8.3 s vs. 79.6 ± 15.9 s, $p < 0.05$). We developed a method to automate the measurement of femorotibial rotation in continuous axial CT scans of patients with osteoarthritis (OA) using a deep-learning approach. This method has the potential to expedite the analysis of patient data in busy clinical settings.

Keywords Deep learning · Automatic measurement · Femorotibial rotation · Computed tomography · Model

✉ Kwang Gi Kim
kimkg@gachon.ac.kr

Sheen-Woo Lee
leesw1@catholic.ac.kr

Gi Pyo Lee
pigyo123@gachon.ac.kr

Jeun Yoon
yoonieunrad@gmail.com

Young Jae Kim
kimyj@gilhospital.com

- ¹ Department of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 1021, Tongil-Ro, Eunpyeong-Gu, Seoul, Republic of Korea
- ² Department of Biomedical Engineering, Gachon University, Seongnam-Si 13120, Gyeonggi-Do, Republic of Korea
- ³ Gachon Biomedical & Convergence Institute, Gachon University Gil Medical Center, Incheon, Republic of Korea
- ⁴ Department of Biomedical Engineering, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

Introduction

Femoral version and tibial torsion angles describe the rotational alignment of the femur and tibia, respectively, defined by the angle between the proximal and distal axes of each bone. These angles are crucial in lower-limb biomechanics and gait; abnormal femoral or tibial rotation can lead to in-toeing or out-toeing gait patterns and is associated with conditions such as femoroacetabular impingement and early osteoarthritis [1–4]. Accurate measurement of femoral version and tibial torsion is essential for assessing the presence and severity of rotational deformities and for guiding clinical decision-making [5].

Conventional assessment of these angles is performed using cross-sectional imaging (typically axial CT). The femoral version is measured as the angle between the femoral neck axis and the distal femoral condylar axis, while the tibial torsion is measured as the angle between the proximal axis and the distal tibial axis through the malleoli [6–8]. This process is usually

done manually: the interpreter identifies the appropriate CT slices for the proximal and distal landmarks of the femur and tibia, draws lines along those landmarks, and then determines the angles between the lines. Such manual measurements are prone to inconsistency—interobserver variability in femoral version has been reported to be as high as 15° and indicates a significant potential for error [9–13].

A further challenge is the lack of a standardized measurement technique. Numerous methods and anatomical landmarks have been proposed for measuring femoral version (Supplementary Table 1), such as using a line through the center of the femoral head and neck, a line along the axis of the femoral neck, or referencing the greater trochanter or femoral shaft [14–16]. No single method has proven clearly superior, and even the tibial torsion angle (for which fewer methods exist) can yield widely varying values—one study reported up to a 100% difference in measured torsion between two techniques, underscoring the methodological inconsistencies in the literature. In addition, the manual workflow is time-consuming. Because of anatomical variability and differences in patient positioning, the relevant proximal and distal landmarks rarely appear on the same CT slice [17]. The operator must scroll through the image stack and repeat the measurement steps on multiple slices to capture all necessary landmarks. Consequently, obtaining bilateral femoral version and tibial torsion measurements for a single patient can take on the order of 10 min, depending on the complexity of the anatomy and the examiner's experience [12, 18–20]. Given these limitations, a more consistent and efficient approach to measuring femoral and tibial rotation is needed.

Deep-learning methods for measurement have been actively embraced by the orthopedic community, such as Cobb's angle [21], pelvic tilt [22], shoulder critical angle [23], and leg length [24]. To our knowledge, however, no study to date has applied a deep-learning technique to measure femoral version and tibial torsion on full-length axial CT scans of the lower extremities. We therefore aimed to develop and validate a deep-learning-based algorithm that can automatically identify the key anatomical landmarks and compute the femoral version and tibial torsion angles from lower-extremity CT images. We hope that this automated approach will improve measurement consistency and significantly reduce the time required for analysis in routine clinical practice.

Material and Methods

Data Collection

This study was approved by our Institutional Review Board (No. PC21RIS10037). Patient consent was waived because

of the retrospective nature of the study. All methods were performed in accordance with the guidelines of the Declaration of Helsinki.

We included adult patient lower-extremity CT scans from a single institution from February 2021 to August 2022, obtained from a PACS database search. Patients with incomplete coverage of the lower extremities or significant imaging artifacts, such as severe motion, were excluded.

Image Acquisition

Examinations were performed using a multidetector CT scanner (Siemens Healthineers Somatom Force; Siemens, Erlangen, Germany). Automated tube current modulation (CARE Dose4D, Siemens Healthineers, Germany) and automated tube voltage selection (CARE, kV) were used for all regions. Patients were scanned in the routine supine position with their knees fully extended. Axial sections were reconstructed at 2-mm thickness. Images with 512×512 resolution were utilized for further processing.

Image Preprocessing

The CT table region was cropped from the images, and two preprocessing steps were performed to improve the segmentation model performance: contrast flattening, using contrast-limited adaptive histogram equalization (CLAHE) with a clip limit of 1.0 and a tile grid size of 16×16 to improve distinction of tissue boundaries. We then followed the RGB superposition technique described by Lee et al. [25], mapping the region-of-interest slice to the green channel, the immediately preceding slice to the red channel, and the immediately succeeding slice to the blue channel, and merged them into a single three-channel image to provide volumetric context to the segmentation model (Fig. 1). By super-positioning these three channels as input, the network can more effectively learn interslice spatial relationships, improving boundary consistency and overall segmentation accuracy.

Data Labeling

Manual Annotation for Bone Segmentation

For segmentation of bone regions in the CT images, we generated labeled data for use in deep-learning training. We used the commercially available labeling software AVIEW (v1.1.44.17, Coreline Soft, Seoul, Korea), manually labeling polygon types that fit the bone regions on each slice. The region of interest (ROI) was drawn on the bone contours of the femur, tibia, and fibula in the CT images. The labeling data are a binary mask image, as shown in Fig. 2, which serves as the ground truth for the training phase of

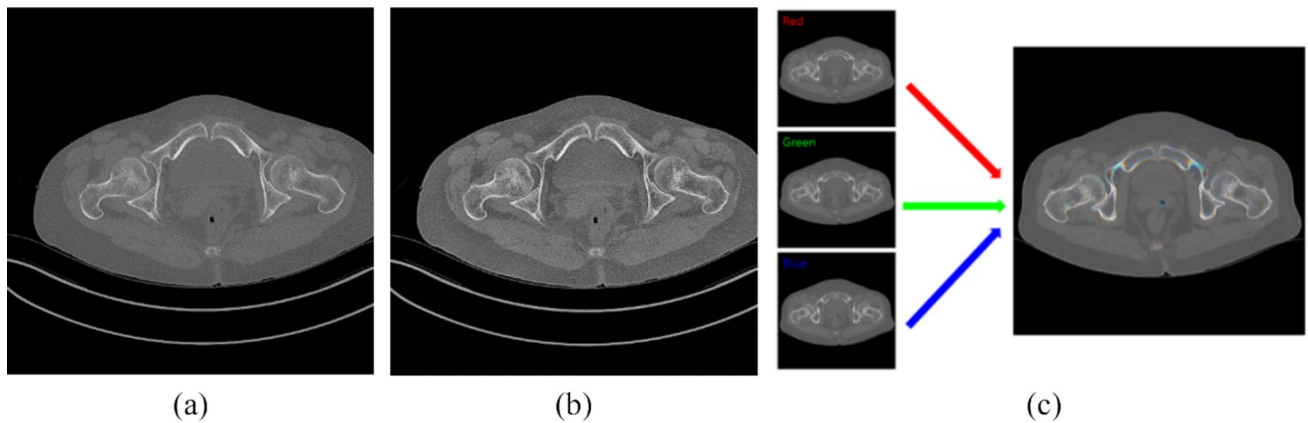


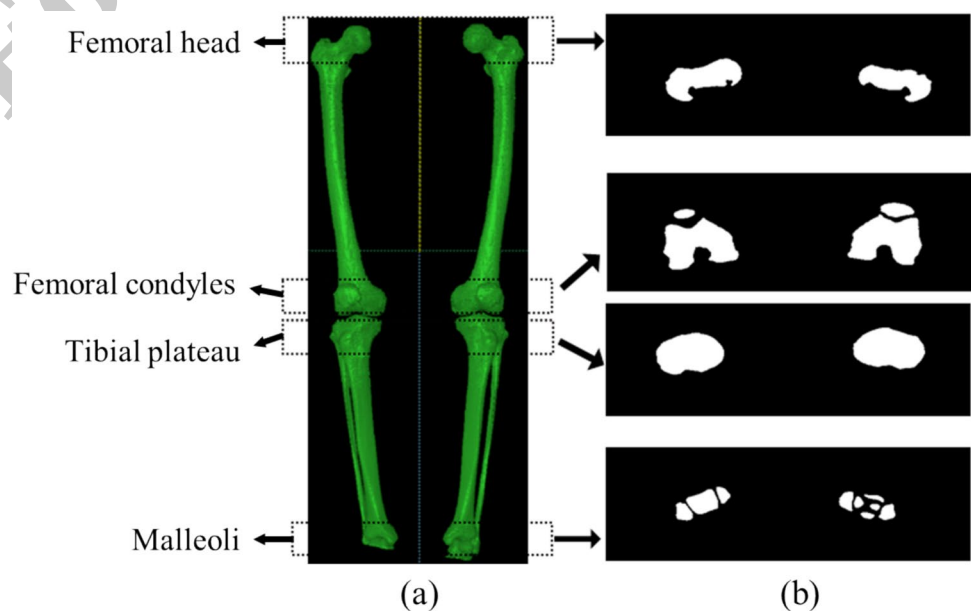
Fig. 1 Sample of histogram equalization and RGB superposition. **a** Original data. **b** Histogram equalization. **c** RGB superposition

the model. Figure 2 a shows the 3D rendered CT image of the lower extremity, covering the femoral head to the tibial malleolar level. The green color in the 3D images indicates the ROI drawn on the bone contours for segmentation. The four dotted boxes from top to bottom in Fig. 2a indicate the locations of the landmarks. Figure 2 b shows the automatic segmentation results of the 2D axial image for each of the four boxes. The images were divided into right and left images. Data labeling was conducted manually by trained operators. All labeling results were subsequently reviewed by a musculoskeletal radiologist, and the reference standard was established through a process of multiple rounds of validation and refinement. The labeling was conducted with the specific objective of ensuring the reliability of the study, thereby establishing a reference standard that was subsequently employed for model training and validation.

Reference Standards by Musculoskeletal Radiologists

Two fellowship-trained musculoskeletal radiologists with 17 and 3 years of subspecialty experience manually assessed the femoral and tibial angles blinded to the patients' information or the AI-obtained results, using commercially available DICOM-viewing programs (TaeYoung Soft, Seoul, Korea, AVIEW v1.1.44.17; Coreline Soft, Seoul, Korea, and 3D Slicer v5.6.2; <https://www.slicer.org/>, USA [26]). The measurement was repeated by a senior radiologist after a washout period of 8 months. The landmarks for measurements were defined following the methods described above. The mean value from the repeated measurements was utilized as the reference standard.

Fig. 2 Examples of manual labeling and part division on lower-extremity CT. **a** The four manually annotated regions, including the femoral head, femoral condyles, tibial plateau, and malleoli, were used as landmark boxes for torsion measurement. **b** Automatic segmentation outputs of the axial 2D slices extracted from each of these boxes



Training

Training involved stepwise automatic segmentation and quantification. First, the femur and tibia are segmented using deep learning. Landmark lines were drawn based on the segmentation, and measurements were performed for the femoral version and tibial torsion. This process is described in detail as follows.

Automatic Bone Segmentation

To segment the regions of the femur and tibia on CT images, we used a deep-learning model based on the 2D Attention U-Net architecture, a widely adopted enhancement of the classic 2D U-Net. In Attention U-Net, feature maps flowing through the encoder–decoder skip connections pass through attention gates, which learn to emphasize relevant regions and suppress background noise. This selective focus on salient features refines the upsampling process, enabling the network to locate anatomical structures more accurately. As a result, Attention U-Net consistently outperforms standard U-Net in medical image segmentation, particularly when distinguishing complex or low-contrast boundaries. The training data set comprised 32 patients, who were initially collected. These patients were randomly assigned into two distinct datasets: a training dataset comprising 24,482 images from 25 patients and a validation dataset comprising 6908 images from 7 patients.

To ensure satisfactory reliability of the verification of the training and measurement algorithms, 238 patients were included in the test dataset. These data were independently collected and were not included in the training and validation datasets. The training data were provided to the model at their native resolution of 512×512 pixels. By foregoing any resizing operations, we preserved the original image information and thus minimized errors in automated angle measurements that can arise from interpolation.

We trained the Attention U-Net model using the Adam optimizer with a learning rate of 1×10^{-4} to minimize the Dice coefficient loss. Training was conducted with a batch size of four for up to 100 epochs, and early stopping was employed to prevent overfitting, with training halted if the validation loss failed to improve for 15 consecutive epochs. In our experiments, early stopping was triggered at epoch 29. Finally, we applied morphological closing and opening operations to the segmented femoral and tibial regions to remove residual background noise and minimize errors in angle measurement.

The processed images were divided into left and right halves to create images of the left and right legs, respectively, for each patient. Then, each leg image was divided into three sections (upper, middle, and lower) based on the

CT image slice numbers. This division was used to examine the axis of angle measurement within each section.

Measuring Femoral Version Angles

The femoral neck axis and bicondylar lines were obtained to measure the femoral version angle. In this study, the reference for the femoral neck axis was based on the method of Reikeras et al., which is a line drawn between the centers of the femoral head and neck on a slice where the posterior and anterior cortices of the neck run parallel to each other. A bicondylar line was defined as a straight line running through the posterior condyle's most prominent portion [7, 27, 28]. To determine the femoral head circumference, we divided the full-leg CT volume into three parts and selected the upper part, which includes the pelvic region. Within this part, we performed bone segmentation on every axial slice, fitted a circumscribed circle to each segmented bone region, and identified the slice with the smallest area difference between the circle and the bone mask as the femoral head reference. And we calculated the following steps: segmentation of the bone area, fitting of a circumscribed circle, and selection of the slice showing the minimal difference between the circle and the bone region.

Along the z -axis in the foot-ward direction, we examined fifty axial slices in the upper segment of the volume, beginning with the first appearance of the femur. In each slice, we identified the largest bone area and drew a circumscribed circle around it. The femoral head was defined as the region where the area difference between the drawn circle and bone region was minimal, indicating the closest match to the circular shape shown in Fig. 3a. To establish the axis of the femoral neck, the width and height of a rectangle (Fig. 3b) circumscribing the bone in each of the 50 previously determined slices were selected. The bone segment with the smallest difference between the width of the drawn rectangle and that of the femur was the femoral neck.

A ROI was detected in the femoral neck using the distance between the outline and internal coordinate values, and a vertical line was drawn at 35% of the long side of the circumscribed rectangle to locate the point where the line met the morphological skeletal outline (Fig. 3c). Figure 3d shows the femoral neck axis as a straight line passing through the two points.

A bicondylar line was obtained from the middle of the three sliced sections (Fig. 4). From the slice in which the patella appeared in the image to the slice in which the condyles disappeared, the slice with the greatest condyle was chosen and used to detect the bicondylar lines. A line was drawn on the condyle of the selected slice between the points with the lowest and highest x values. The drawn line moved parallel to the downward (positive direction along the y -axis) direction to make tangential contact with the inferior border

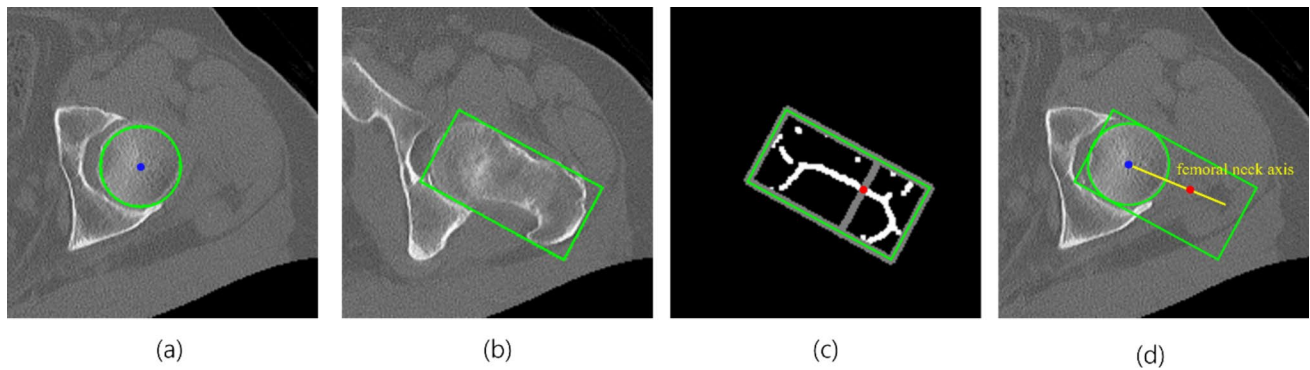


Fig. 3 Process of detecting the femoral neck axis. **a** Femoral head identification. **b** Femoral neck localization. **c** Detection in the femoral neck (red dot: intersection of this line with the bone contour defined

the neck coordinate). **d** Overlay of the detected femoral neck axis (yellow line) on the original axial CT slice

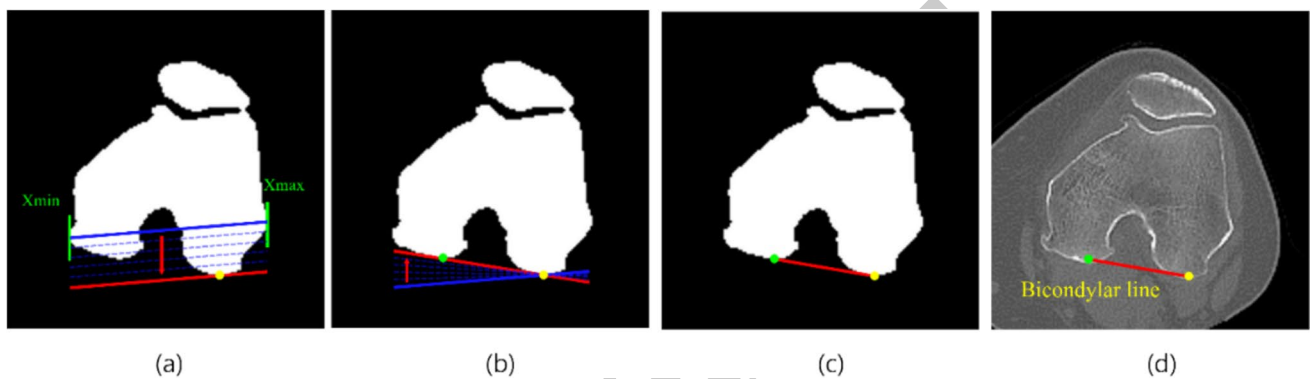


Fig. 4 Process of detecting the bicondylar line. **a** Binary mask of the condylar region extracted from the middle section. The x -min and x -max points (green brackets) define the initial horizontal span. **b** Initial line (red) drawn between the lowest and highest x -value points

on the condyle mask. **c** The line is translated downward and its slope adjusted so that it becomes tangent to the inferior border of the condyle mask (blue), identifying the final bicondylar line (red). **d** Overlay of the detected bicondylar line on the original axial CT slice

of the condyle. The tangent was calculated by changing the slope of the straight line by moving the point furthest from the tangent upward (in a negative direction on the x -axis) in the image, either side of the two points that make up the straight line that was moved to be tangent to the condyles. The resulting tangent line is labeled as the bicondylar line.

The femoral version angle was defined as the angle between the bicondylar line and femoral neck axis (Fig. 5).

Measuring the Tibial Torsion Angles

The proximal tibial and distal bimalleolar lines were acquired. The proximal tibial line spans the most prominent posterior margin of the tibial plateau, whereas the bimalleolar line traverses the midpoint between the medial and lateral malleoli.

To acquire the proximal tibial line, middle slices were taken from the three cut regions, and the search was undertaken in reverse, beginning with the slice where the tibial plateau first emerged and ending with the slice where the

patella ended. A slice of the tibial plateau with the largest dimension was selected, and a line was drawn connecting the two points with the lowest and highest x -coordinate values (Fig. 6a). To obtain the bottom portion of the tibial plateau, a temporary line was moved down the image (in the direction where the y -coordinate value was positive) (Fig. 6a). To obtain the bottom portion of the tibial plateau, the temporary line was then moved down the image (in the direction where the y coordinate value is positive) (Fig. 6b). The opposite endpoint of the tangent line was elevated (in the direction where the y -value of the coordinates became negative) until it intersected the tibial plateau region at their two spots (Fig. 6c). The resulting tangent line was the proximal tibial line (Fig. 6).

The bimalleolar line was determined by assessing the ankle bones, including the malleoli, from the last region of the three divided slices. The images were searched for three or more bone sections in the slices, and the center of gravity for each bone region was identified (Fig. 7a). The distance between all the center of gravity points was

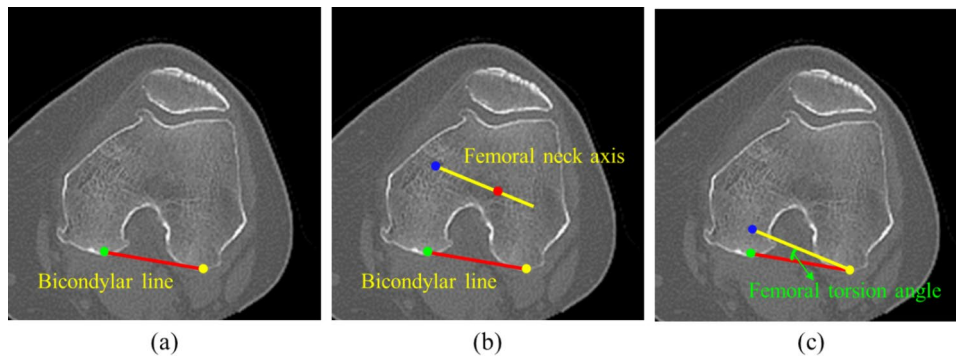


Fig. 5 Process of automatic femoral version angle measurement. **a** Bicondylar line detection on the axial CT slice, drawn between the most inferior points of the medial (green dot) and lateral (red dot) condyles. **b** Femoral neck axis determination, shown as a line (yellow) connecting the center of the femoral head (blue dot) and the

detected neck coordinate (red dot). **c** Computation of the femoral torsion angle (green arc), defined as the acute angle between the femoral neck axis (yellow) and the bicondylar line (red) on the same axial slice

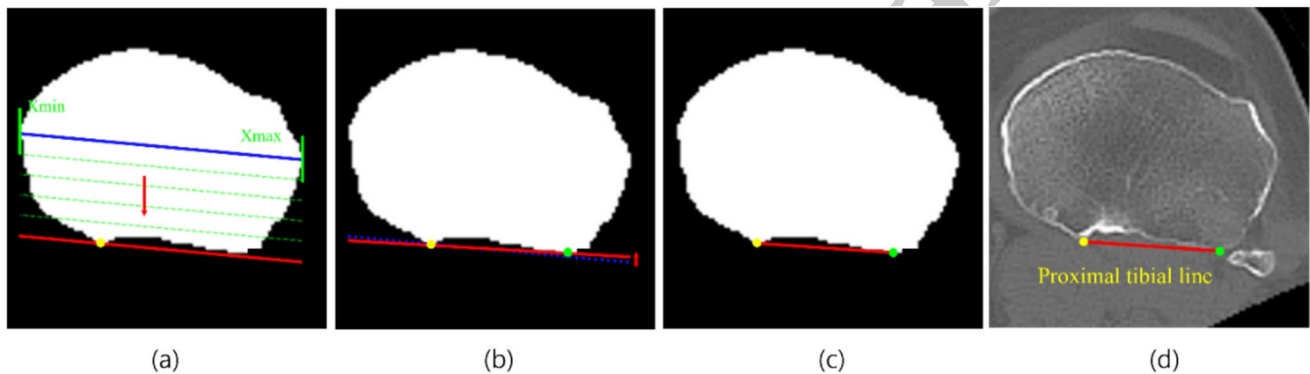


Fig. 6 Process to find proximal tibial line. **a** Draw a provisional line between the lowest and highest x -coordinate points on the tibial plateau mask. **b** Translate the line downward to approximate the inferior

border. **c** Adjust the opposite endpoint upward until the line is tangent at two points. **d** Overlay the final proximal tibial line on the axial CT slice

calculated (Fig. 7b), and the line representing the longest distance was determined for each slice (Fig. 7c). The bimalleolar line was the longest straight line for all slices (Fig. 7d). The tibial torsion angle was then established by measuring the angle between the proximal tibial and bimalleolar lines, as shown in Fig. 8. The process is illustrated in Figs. 7 and 8 [29].

Test Set

After establishing the algorithm, different sets of lower-extremity CT scans of adult patients were searched using PACS, and the anonymized data were used to validate the automated CT segmentation. The algorithm-derived angles were compared to those measured by a musculoskeletal radiologist.

Acquisition Time

The time for two musculoskeletal radiologists to measure the femoral version and tibial torsion was recorded in ten randomly selected studies from the test set. The acquisition times for the right and left sides were compared with those of the algorithm.

Study Environment

All development and timing experiments were performed on a workstation running a 64-bit operating system with Python 3.7.15. The system was equipped with an Intel® Xeon® Gold 5218R CPU 2.10 GHz (Intel, Santa Clara, CA, USA) and 256 GB of RAM. Deep-learning computations were executed using TensorFlow 1.15.0 and Keras 2.3.1 on a 24 GB

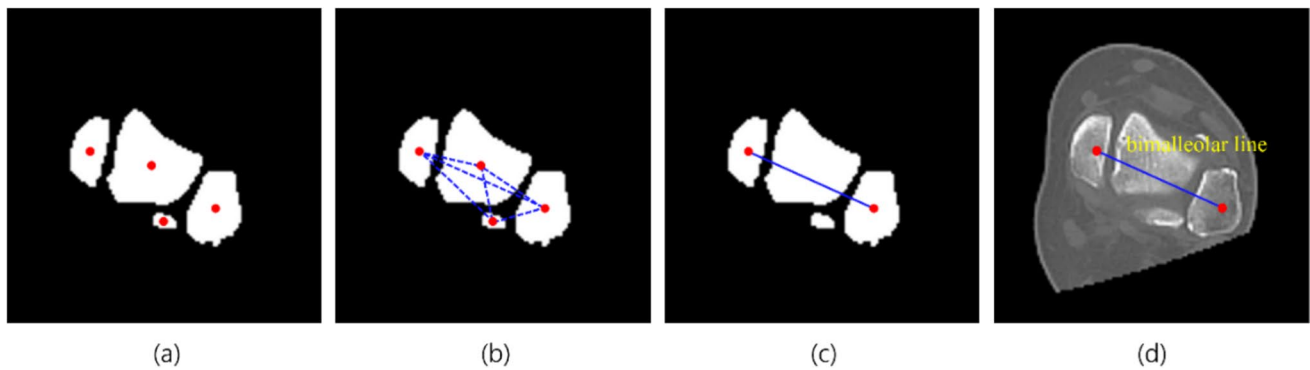


Fig. 7 Process of detecting the bimalleolar line. **a** Calculate the centroids (red dots) of each malleolar bone region in the distal section masks. **b** Compute all pairwise distances between centroids (dashed blue lines). **c** Select the longest centroid-to-centroid line in each slice

as the bimalleolar candidate (solid blue line). **d** Designate the overall longest line across all slices as the final bimalleolar line, overlaid here on the axial CT image

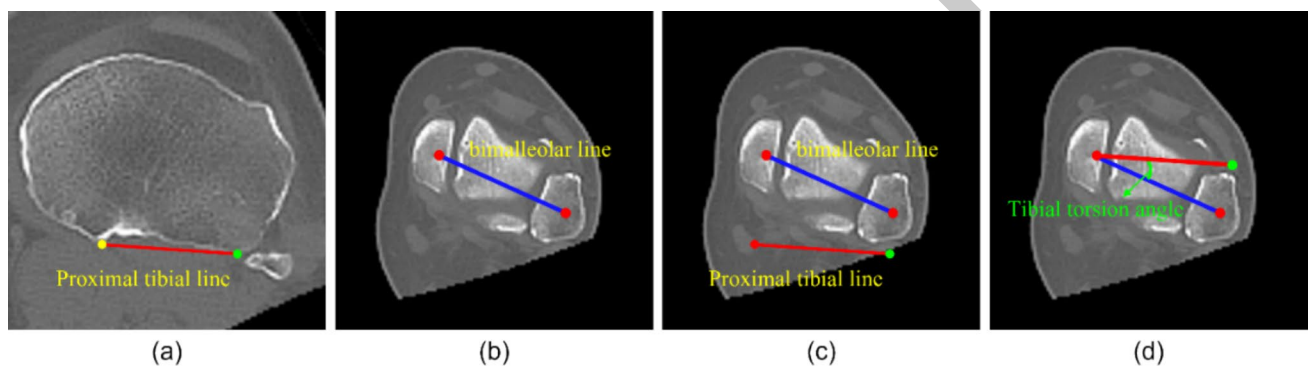


Fig. 8 Process of automatic tibial torsion angle measurement. **a** Proximal tibial line (red) drawn on the axial CT slice at the tibial plateau level. **b** Bimalleolar line (blue) detected on the distal axial slice connecting the centroids of the medial and lateral malleoli. **c** Overlay of

the proximal tibial line (red) and bimalleolar line (blue) on a single slice to illustrate relative orientation. **d** Tibial torsion angle (green arc) measured as the acute angle between the proximal tibial and bimalleolar lines on the same axial image

VRAM of NVIDIA GeForce RTX A5000 GPU (NVIDIA, Santa Clara, CA, USA).

Statistical Analysis

The results were tested for normal distribution using Kolmogorov–Smirnov analysis. Accuracy, specificity, sensitivity, and Dice similarity coefficient (DSC) were employed as performance metrics for the segmentation model. Additionally, the 95% of Hausdorff distance (HD95) was used to assess boundary accuracy; HD95 represents the distance within which 95% of the boundary points lie, providing a robust measure of contour alignment while reducing the influence of extreme outliers.

The agreement between the femorotibial angles derived from the algorithm and those by the radiologists was analyzed using Spearman's correlation and Bland–Altman analysis. A senior MSK radiologist qualitatively assessed DL-generated landmark lines on CT scans and classified

the errors at each of the four landmark levels as follows: none (0), line on accessory bone such as fabella or fibular head (acc), on completely different level such as metaphysis (dl), on osteophyte (os), deviated due to metallic prosthesis (mp), and severe joint deformity (jt), unclassifiable (un). The results are shown in Figs. 10, 11, and 12 and Table 3.

To check the association of measurement results with gender or age, intra- and inter-reader agreements of radiologists were assessed using the results by two musculoskeletal specialists. Since the patient group was heavily female- and elderly-based, any association between the results and gender or age was analyzed using Mann–Whitney *U* and Spearman's tests. The acquisition times for algorithm- and radiologist-derived results were evaluated using the Wilcoxon signed-rank test. MedCalc (MedCalc Software Ltd, v.23.1.3, Belgium) and SPSS (IBM SPSS Statistics for Windows, v29.0, USA) programs were used for statistical analysis, and the threshold for statistical significance was set at $p < 0.05$.

Results

Patient Characteristics

Complete CT scans of 270 adult patients were included (median age, 69 years; range 63–75 years; female/male ratio, 235:35). To develop the segmentation model via deep learning, 32 patients were randomly selected, with 25 CT scans designated for model training and 7 for validation. A total of 24,482 sliced images and 6908 sliced images from the selected patient cohorts were processed during the training and validation phases, respectively. A total of 238 patients were used to test the model and accuracy of the automatic measurement algorithm.

Segmentation Performance

For the test dataset with 238 patient CT scans of patients, the entire segmentation model prediction yielded an DSC of $93.12 \pm 0.01\%$, a sensitivity of $92.16 \pm 0.02\%$, a specificity of $99.96 \pm <0.01\%$, and a HD95 of 2.14 ± 2.39 . According to Table 1, the segmentation model demonstrated consistently high performance across both legs and all anatomical sections. For whole-leg segmentation, the mean Dice coefficients were $93.59 \pm 0.87\%$ on the left and $93.21 \pm 0.91\%$ on the right, with sensitivities of $92.52 \pm 2.06\%$ and $92.52 \pm 2.03\%$, and specificities exceeding 99.95% on both sides. Boundary accuracy, as measured by HD95, was below 1 mm (0.93 ± 0.49 mm left, 0.91 ± 0.38 mm right).

Sectional analysis revealed the highest Dice scores in the middle regions ($94.44 \pm 0.77\%$ left, $94.40 \pm 0.82\%$ right) and the lowest in the lower regions ($92.17 \pm 1.42\%$ left, $90.94 \pm 1.49\%$ right). Sensitivity followed a similar pattern, peaking in the middle ($94.14 \pm 1.14\%$ left, $92.52 \pm 2.03\%$ right) and dipping in the lower Sects. ($90.29 \pm 5.25\%$ left, $90.06 \pm 5.56\%$ right). HD95 was smallest in the middle (0.75 ± 0.35 mm left, 0.70 ± 0.36 mm right) and largest in the lower regions (1.24 ± 1.16 mm left and 1.24 ± 0.76 mm right), indicating slightly greater boundary deviation at the distal ends.

Representative segmentation outputs present axial CT slices at four anatomical levels, the proximal femoral head, femoral condyles, tibial plateau, and malleoli, in Fig. 9. This results in the maintenance of boundary fidelity, even in low-contrast regions, and the minimization of false positives in adjacent soft tissues.

Measurement of Femorotibial Angles

The angles generated by the algorithm were as follows: right and left femoral version angles, $10.94 \pm 6.96^\circ$ and $9.17 \pm 6.50^\circ$; right and left tibial torsion angles, $23.47 \pm 8.51^\circ$ and $26.19 \pm 9.18^\circ$, respectively (Figs. 10, 11, and 12). The manual results by the musculoskeletal radiologist were as follows: right and left femoral version angles, $15.60 \pm 26.90^\circ$ and $10.25 \pm 15.25^\circ$, respectively; right and left tibial torsion angles, $24.1 \pm 14.27^\circ$ and $25.75 \pm 12.34^\circ$, respectively. The data did not exhibit normal distribution at any levels ($p < 0.05$). The intrareader and interreader correlation coefficients were 0.64 (95% CI 0.644–0.721, $p < 0.0001$) and 0.54 (0.48–0.60, $p < 0.0001$). Table 2 shows the angle results according to gender or age.

No statistically significant differences in results were observed between males and females (all $p > 0.05$). Correlation with age was minimal across all angles ($|r| < 0.15$).

AI-derived angle measurements showed no significant differences from radiologist assessments, supporting their robustness across demographic subgroups. The correlation coefficient for the femoral angle values obtained automatically and manually was 0.70, while for the tibial torsion angle value, it was 0.58. For right femur, left femur, right tibia, and left tibia, the correlation coefficients were 0.73 (95% CI 0.65–0.79, $p < 0.0001$), 0.66 (95% confidence interval [CI]: 0.57–0.73, $p < 0.0001$), 0.59 (0.49–0.68, $p < 0.0001$), 0.57 (0.46–0.66, $p < 0.0001$), respectively, indicating good correlation (Fig. 13).

Bland–Altman analysis was performed to compare radiologist and algorithm measurements of femoral version and tibial torsion angles (Fig. 14). On the abscissa lies the average of the two methods, and on the ordinate the difference between radiologist and algorithm values. The

Table 1 Segmentation performance metrics by direction and algorithmic region

Direction	Part	DSC	Sensitivity	Specificity	HD95
Left	Whole	93.59 ± 0.87	92.52 ± 2.06	99.97 ± 0.01	0.93 ± 0.49
	Upper	94.03 ± 2.73	92.48 ± 3.71	99.98 ± 0.01	0.86 ± 0.99
	Middle	94.44 ± 0.77	94.14 ± 1.14	99.94 ± 0.01	0.75 ± 0.35
	Lower	92.17 ± 1.42	90.29 ± 5.25	99.96 ± 0.02	1.24 ± 1.16
Right	Whole	93.21 ± 0.91	92.52 ± 2.03	99.95 ± 0.01	0.91 ± 0.38
	Upper	94.31 ± 1.66	92.27 ± 3.35	99.98 ± 0.01	0.84 ± 0.73
	Middle	94.40 ± 0.82	92.52 ± 2.03	99.95 ± 0.01	0.70 ± 0.36
	Lower	90.94 ± 1.49	90.06 ± 5.56	99.95 ± 0.02	1.24 ± 0.76

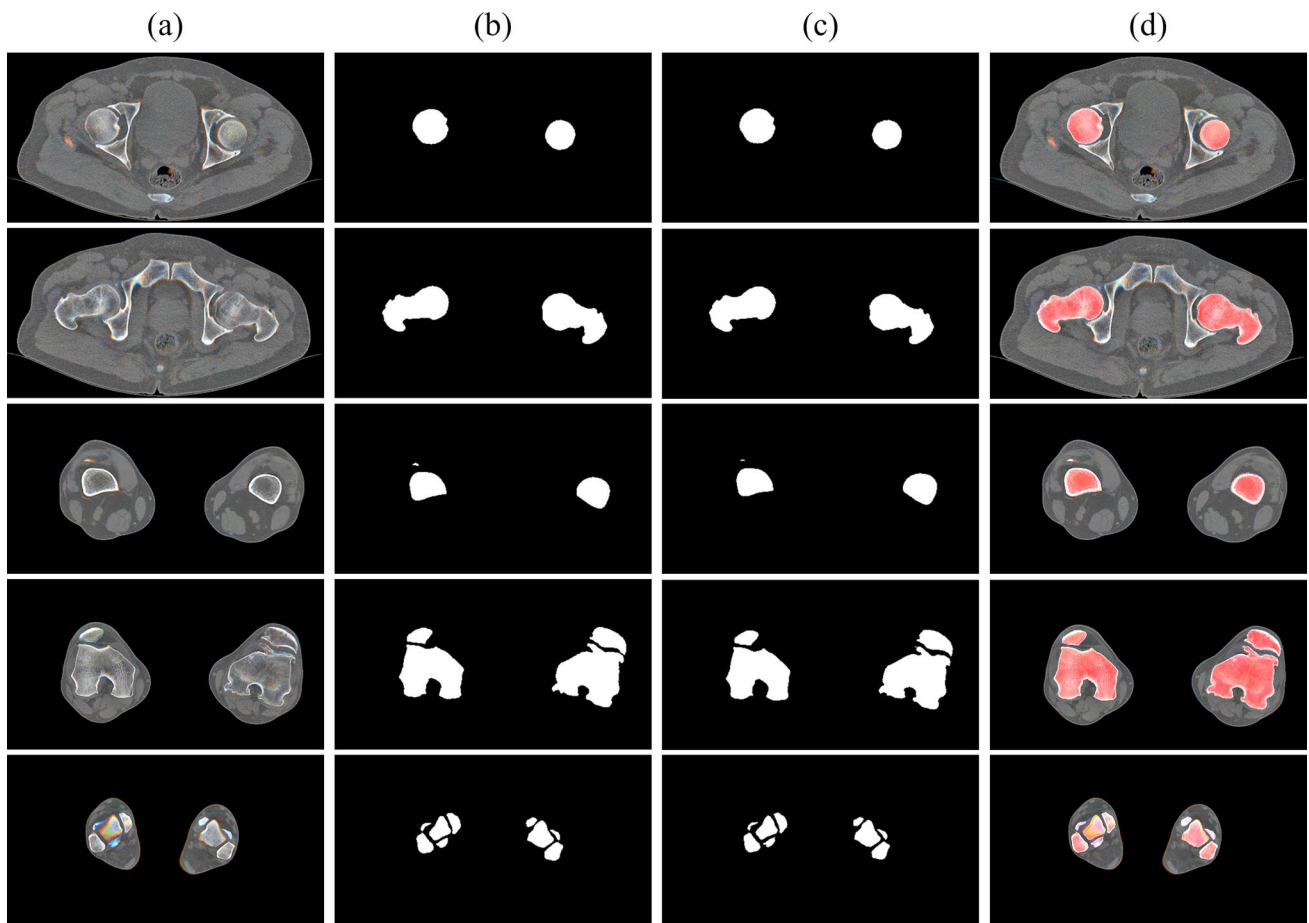


Fig. 9 Segmentation model result. **a** Input: RGB superposition images. **b** Ground truth mask. **c** Predicted image by segmentation model. **d** Overlay of input images and predicted segmentation results

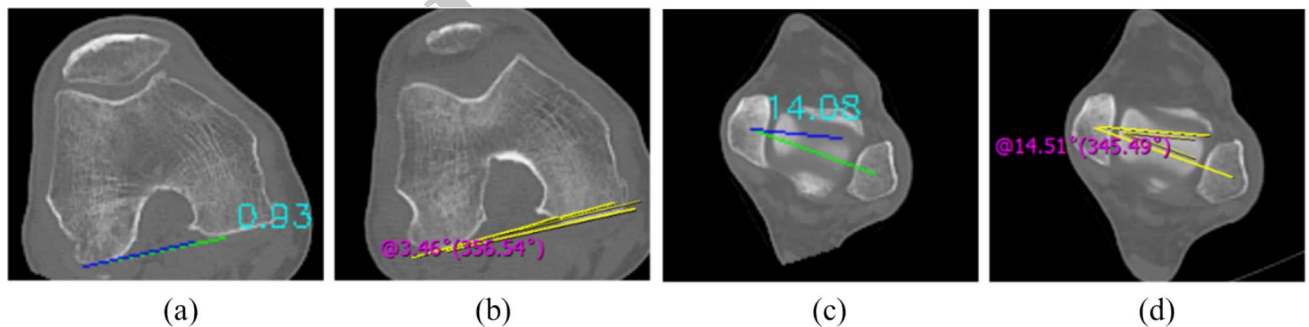


Fig. 10 Femoral angles, algorithm output (**a**, **c**), and manually measured results (**b**, **d**)

mean differences and 95% limits of agreement (LoA) were as follows: left femoral version, 0.8° (LoA -9.6° to 11.2°); right femoral version, -2.6° (LoA -12.9° to 7.7°); left tibial torsion, 2.2° (LoA -12.3° to 16.8°); and right tibial torsion, -1.8° (LoA -16.0° to 12.4°). All biases were small, and every LoA lay within the a priori-defined clinical

tolerance thresholds, demonstrating that the radiologist and algorithm measurements are in good agreement.

When errors were qualitatively analyzed for each case, the right distal femur and left proximal tibia had the most errors ($p < 0.001$) (Table 3). There was no statistically significant association between the location and error types

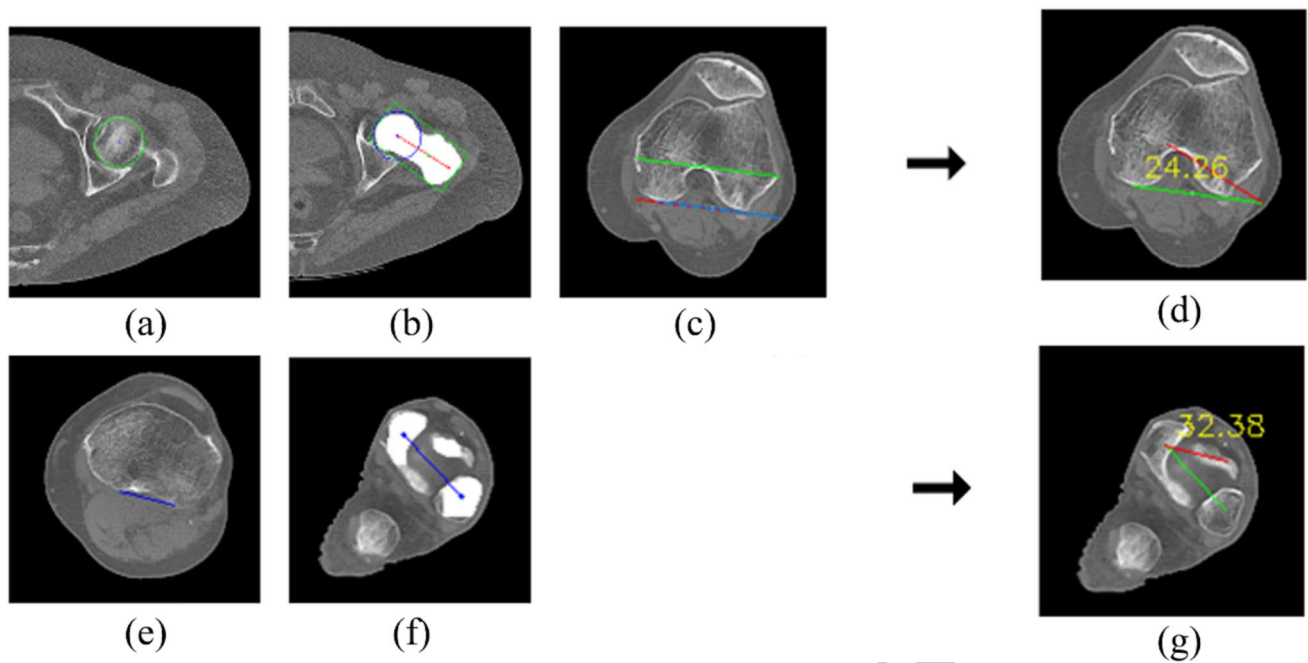


Fig. 11 Example of good measurement by the algorithm

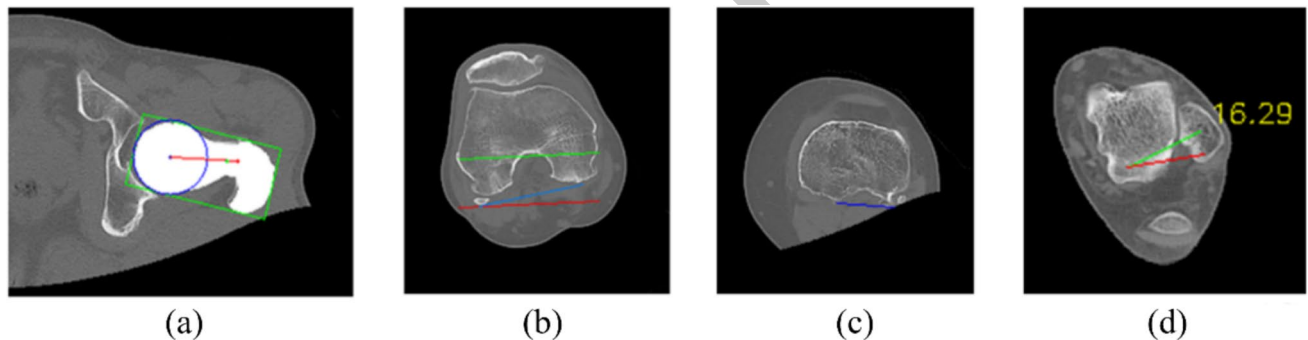


Fig. 12 Examples of erroneous labeling. **a** shows that the algorithm yielded a line matching the Lee method and not the Reikeras method that we trained. **b** The fabella is incorrectly

included in the posterior femoral margin. **c** Osteophytes from the posterior tibia are included in osteoarthritic patients. **d** The medial malleolus was not recognized by the algorithm

($p = 0.88$). The acquisition time for the radiologist to measure the femoral version and tibial torsion was 79.6 ± 15.9 (75.7 [68.4–92.9]) seconds per leg. That of the DL (deep learning) was 57.5 ± 8.3 (60 [51.6–64.2]) seconds per leg ($p < 0.05$).

Discussion

We developed and validated a deep-learning-based method for automatic measurement of the femorotibial rotational angle using lower-extremity CT. Using an Attention U-Net architecture for bone segmentation, our model demonstrated

good accuracy and strong agreement with expert measurements. By automating the identification of anatomical landmarks on full-length CT images, this approach can significantly streamline the workflow for assessing femoral version and tibial torsion, potentially reducing the measurement time from minutes to seconds and improving consistency across observers.

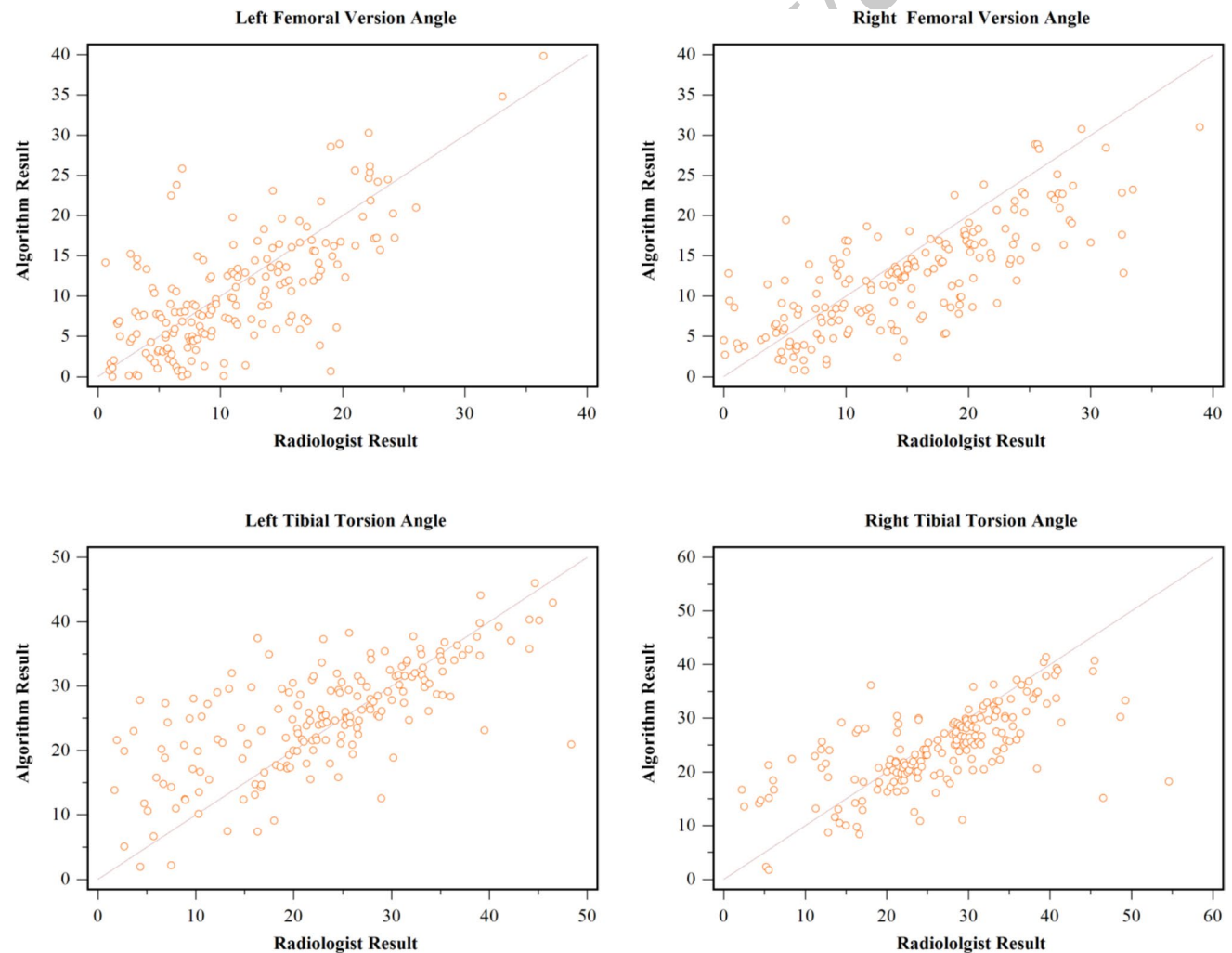
To the best of our knowledge, no prior study has utilized deep learning to measure femorotibial angles in continuous axial CT scans of the lower extremities. The application of deep learning in medical imaging has expanded to include various diagnostic applications such as lesion detection and triage. The deep-learning method has been applied

Table 2 Comparison of lower extremity angle measurements by sex and age: AI and radiologist assessment

		Sex			Age		Comparing AI and radiologist	
	Angle (Mean degrees $\pm$ SD)	Male	Female	(<i>p</i> value)	Correlation coefficient	(<i>p</i> -value)	<i>p</i> value (sex)	<i>p</i> -value (age)
Right	AI-femoral	9.19 $\pm$ 5.94	15.63 $\pm$ 8.1	(0.08)	0.001	(0.99)	0.7	0.74
	Rad-femoral	9.0 $\pm$ 4.7	12.72 $\pm$ 6.69	(0.06)	0.01	(0.88)		
Left	AI-femoral	6.94 $\pm$ 4.3	11.0 $\pm$ 7.62	(0.07)	0.14	(0.05)	0.52	0.23
	Rad-femoral	9.52 $\pm$ 5.53	11.58 $\pm$ 6.89	(0.24)	0.07	(0.38)		
Right	AI-tibial	25.11 $\pm$ 10.33	26.4 $\pm$ 9.7	(0.41)	0.09	(0.2)	0.74	0.58
	Rad-tibial	25.4 $\pm$ 7.55	24.32 $\pm$ 7.46	(0.55)	−0.04	(0.63)		
Left	AI-tibial	22.12 $\pm$ 11.13	23.18 $\pm$ 10.24	(0.6)	−0.01	(0.91)	0.76	0.15
	Rad-tibial	26.62 $\pm$ 7.19	25.12 $\pm$ 8.49	(0.17)	0.03	(0.65)		

Values are presented as mean $\pm$ standard deviation. No statistically significant differences were observed between male and female patients (all $p > 0.05$). Correlation with age was minimal across all angles ($|r| < 0.15$). AI-derived angle measurements showed no significant differences from radiologist assessments, supporting their robustness across demographic subgroups

AI-femoral AI-measured femoral version angle, *Rad-femoral* radiologist-measured femoral version angle, *AI-tibial* AI-measured tibial torsion angle, *Rad-tibial* radiologist-measured tibial torsion angle)

**Fig. 13** Scatterplot for the correlation between the results from algorithm and radiologist

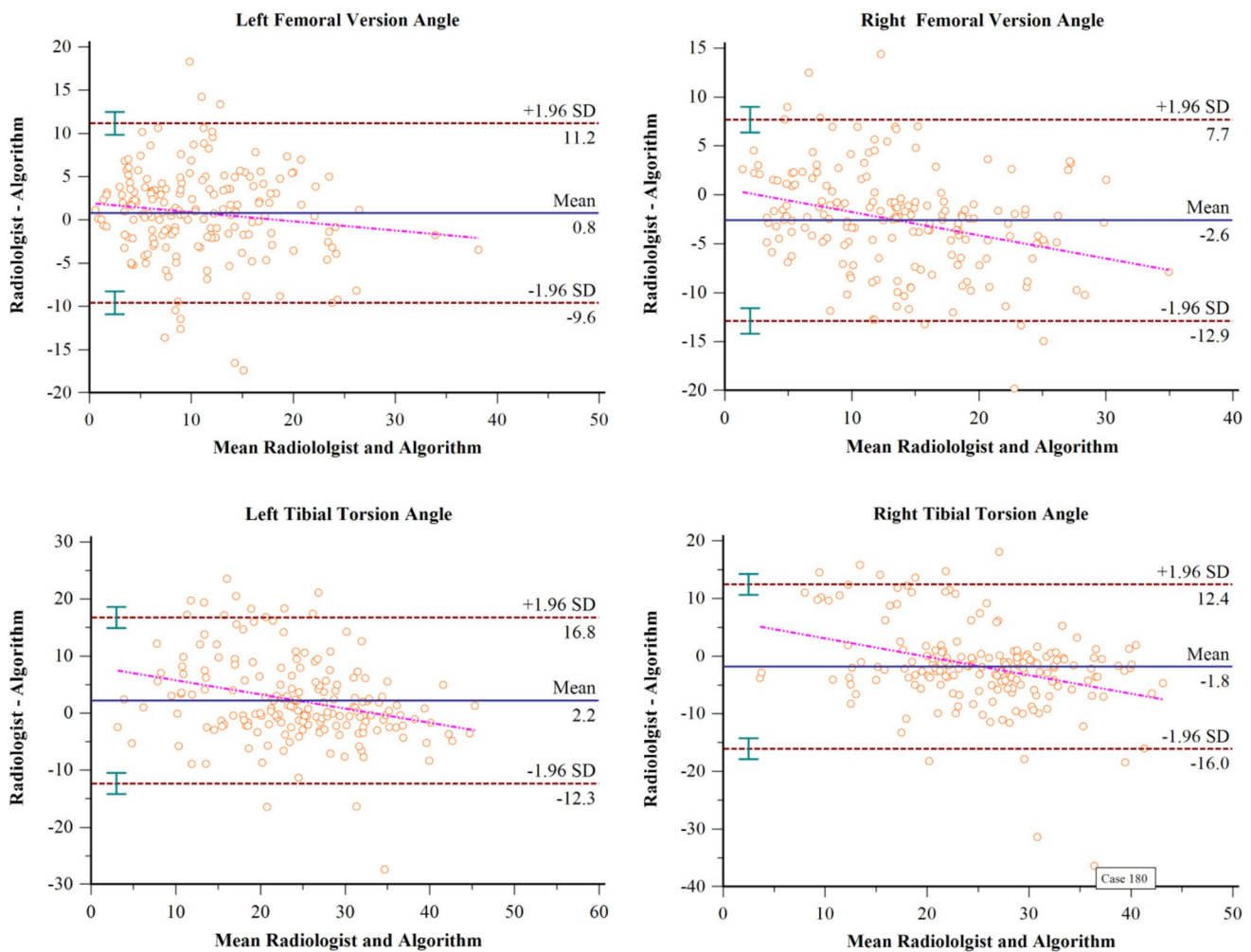


Fig. 14 Bland–Altman plot of the agreement between the algorithm and radiologist

Table 3 The location and types of errors in landmark line generation by DL

Location /Error type	0		acc		dl		jt		mp		un		os		Total number of errors
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%	
1	235	98.74	0	0	2	0.84	0	0	1	0.42	0	0	0	0	3
2	193	81.09	9	3.78	13	5.46	0	0	6	2.52	7	2.94	10	4.2	45
3	157	65.97	38	15.97	14	5.88	0	0	0	0	9	3.78	20	8.4	81
4	232	97.48	0	0	4	1.68	0	0	0	0	2	0.84	0	0	6
5	228	95.8	0	0	9	3.78	0	0	1	0.42	0	0	0	0	10
6	185	77.73	20	8.4	18	7.56	0	0	6	2.52	4	1.68	5	2.1	53
7	152	63.87	52	21.85	14	5.88	0	0	1	0.42	0	0	19	7.98	86
8	226	94.96	0	0	7	2.94	1	0.42	0	0	4	1.68	0	0	12
Total	1608		119		81		1		15		26		54		

Error types in four landmark locations on each leg were classified as follows: none (0), erroneous line on osteophyte (os), on accessory bone such as fabella or fibular head (acc), on completely different level such as metaphysis (dl), deviated due to metallic prosthesis (mp), and severe joint deformity (jt), and unclassifiable (un). Location 1 right proximal femur, 2 right distal femur, 3 right proximal tibia, 4 right distal tibia, 5 left proximal femur, 6 left distal femur, 7 left proximal tibia, 8 left distal tibia. Location 3 and 7 (right femur distal and left tibia proximal) had the most errors ($p < 0.001$). There was no statistically significant association between the location and error types ($p = 0.88$).

to measure other orthopedic parameters such as Cobb's angle, pelvic tilt, shoulder critical angle, and bone age [30]. Our work extends this concept to 3D volumetric imaging, focusing on continuous CT, which is routinely used during orthopedic practice, tackling the complex lower-extremity anatomy. Despite this complexity, the algorithm's results in our study were comparable to those of an expert radiologist, while offering a substantially faster analysis.

It is also important to consider research using other imaging modalities. Schock et al. used lower leg MRI to assess femoral and tibial torsion [19]. However, MRI is not routinely used for full lower-extremity torsional evaluation due to longer scan times, higher cost, and limited availability compared to CT. In current clinical practice, CT remains the standard for evaluating rotational deformities in adults. Our CT-based deep-learning algorithm is therefore well-aligned with existing workflows and could be readily adopted to provide automatic femorotibial angle measurements in a practical, time-efficient manner.

Our study has several limitations. First, inaccuracies in landmark positioning by osteophytes or accessory bones meant that patients with severe deformities or arthritic changes required label adjustment and secondary verification. In one instance, an incorrect level was identified in a figure without such a deformity. We aim to refine the algorithm by augmenting the case volume and including additional patients with deformities. Second, the composition of our patient population was heavily skewed to the female elderly. Elderly female is known to be much more prone to painful osteoarthritis of the knee [31]. It is crucial to accurately evaluate limb malrotation prior to surgery in severely arthritic joints, so we believe that it is important to mirror the epidemiology of knee osteoarthritis. Thirdly, we excluded the adolescent group, where abnormal torsional alignment of the lower extremity is a common reason for hospital visits. Considering the radiation exposure, pediatric patients may benefit from modalities such as MRI for cross-sectional assessments. We hope to expand our automated algorithm in the future to incorporate data from a younger patient demographic. Fourth, several images by DL involved different methods of drawing the femoral neck line compared with the original training, but the methods were within the known variations of femoral measurements and thus were not classified as errors. The femoral and tibial angles are subject to wide interobserver and intraobserver variations, as mentioned before, and depend on the readers' expertise, utilized landmark, and software. But improvement in the performance of the algorithm is necessary, and to this end, it is imperative that the bone segmentation step of our AI model attain more precise boundary delineation. We plan to augment our training by collecting additional patient data at our institution and through collaborations with other medical centers. We will then use transfer learning on the expanded

dataset to improve model accuracy. In addition, further research will be conducted to implement 3D CNN-based segmentation architectures. The objective of this implementation is to fully exploit the volumetric information presented in the complete CT data.

Interestingly, although the quantitative parameter for DL performance indicated better segmentation at mid-levels of the lower leg, such as around the knee, the qualitative assessment by the radiologists indicated otherwise, i.e., the distal femur and proximal tibia exhibited more errors. The metrics that we used such as accuracy, Dice, and HD95 scores are commonly accepted methods for the evaluation of segmentation. But recent papers have voiced difficulties in applying such metrics for subtle tissue contrast or overlapping structures, the answer for which is yet to be addressed [32]. It would be an important research topic in the future to develop better strategies for DL validation.

Lastly, there were substantial differences in the right and left femur and between genders, which may be perceived as erroneous; however, research indicates that the left tends to show increased anteversion in the healthy population [33], consistent with our result. Additional application in diverse patient groups will be essential to enhance the consistency of the results.

In conclusion, we developed a method to automate femorotibial rotation from continuous axial CT scans of patients with osteoarthritis (OA) using a deep-learning approach. We hope that this approach will aid the rapid analysis of patient data in busy clinical settings.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-025-01641-0>.

Acknowledgements We would like to thank Editage (www.editage.co.kr) for English language editing.

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Sheen-Woo Lee, Gi Pyo Lee, Jeun Yoon, and Young Jae Kim. The first draft of the manuscript was written by Sheen-Woo Lee and Gi Pyo Lee, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by the Nuri-light Radiological Medicine Research Society and also supported by the Korea Medical Device Development Fund grant funded by the Korean government (Ministry of Science and ICT, Ministry of Trade, Industry and Energy, Ministry of Health & Welfare, and Ministry of Food and Drug Safety) (Project Number: RS-2022-00141526) and by the Technology Innovation Program (RS-2025-02305698, Development of On-Device AI Digital Medical Products Utilizing Synthetic Data and Synthetic Technology for Atypical Medical Data) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Declarations

Ethical Approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the

Institutional Review Board of Eunpyeong St. Mary's Hospital (No. PC21RISI0037).

Consent to Participate This study was conducted using retrospectively collected data. Written informed consent was waived due to the use of anonymized CT images with no identifiable personal information. This study was approved by the Institutional Review Board of Eunpyeong St. Mary's Hospital (IRB No. PC21RISI0037).

Consent for Publication The authors affirm that human research participants provided informed consent for publication of the images in Figure(s).

Conflict of Interest The authors declare no competing interests.

Disclosure This manuscript has not been published previously. It is not under consideration for publication elsewhere, and its publication was approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out. If accepted, the manuscript will not be published elsewhere in the same form, in English, or in any other language, including electronically, without the written consent of the copyright holder.

References

1. Y.W. Chae, S. Park, J.W. Park, Pelvic, Hip, and Knee Kinematics of Stair Climbing in People with Genu Varum, *The Korean Society of Physical Therapy*, 2018, pp. 14–22.
2. D. Goutallier, S. Van Driessche, O. Manicom, E. Sariali, J. Bernageau, C. Radier, Influence of lower-limb torsion on long-term outcomes of tibial valgus osteotomy for medial compartment knee osteoarthritis, *J Bone Joint Surg Am*, 88 (2006) 2439–2447.
3. J.A. Gruskay, A.T. Fragomen, S.R. Rozbruch, Idiopathic Rotational Abnormalities of the Lower Extremities in Children and Adults, *JBJS Rev*, 7 (2019) e3.
4. D. Tonnis, A. Heinecke, Acetabular and femoral anteversion: relationship with osteoarthritis of the hip, *J Bone Joint Surg Am*, 81 (1999) 1747–1770.
5. B. Van Fraeyenhove, J.C.F. Verhaegen, J. Grammens, G. Mestach, E. Audenaert, A. Van Haver, P. Verdonk, The quest for optimal femoral torsion angle measurements: a comparative advanced 3D study defining the femoral neck axis, *J Exp Orthop*, 10 (2023) 141.
6. F. Berryman, P. Pynsent, C. McBryde, A semi-automated method for measuring femoral shape to derive version and its comparison with existing methods, *Int J Numer Method Biomed Eng*, 30 (2014) 1314–1325.
7. O. Reikeras, A. Hoiseth, Torsion of the leg determined by computed tomography, *Acta Orthop Scand*, 60 (1989) 330–333.
8. W. Waldt, *Measurements and Classifications in Musculoskeletal Lower Limb Alignment*, Thieme Medical, 2023.
9. A.J. Bizdikian, A. Assi, Z. Bakouny, F. Yared, E. Saghbini, G.E. Bakhos, S. Esber, N. Khalil, J. Otayek, J. Ghanimeh, C. Sauret, W. Skalli, I. Ghanem, Validity and reliability of different techniques of neck-shaft angle measurement, *Clin Radiol*, 73 (2018) 984 e981–984 e989.
10. F. Schmaranzer, M.K. Meier, T.D. Lerch, A. Hecker, S.D. Step-pacher, E.N. Novais, A.M. Kiapour, Coxa valga and antetorta increases differences among different femoral version measurements: potential implications for derotational femoral osteotomy planning, *Bone Jt Open*, 3 (2022) 759–766.
11. M. Scorcelletti, N.D. Reeves, J. Rittweger, A. Ireland, Femoral anteversion: significance and measurement, *J Anat*, 237 (2020) 811–826.
12. J.M. Stephen, R.A. Teitge, A. Williams, J.D.F. Calder, H. El Daou, A Validated, Automated, 3-Dimensional Method to Reliably Measure Tibial Torsion, *Am J Sports Med*, 49 (2021) 747–756.
13. J. Yoshitani, K.H. Sunil Kumar, S. Ekhtiari, V. Khanduja, The conundrum in the measurement of femoral anteversion for young adults with hip pathology, *Bone Joint J*, 105-B (2023) 1239–1243.
14. S.B. Murphy, S.R. Simon, P.K. Kijewski, R.H. Wilkinson, N.T. Griscom, Femoral anteversion, *J Bone Joint Surg Am*, 69 (1987) 1169–1176.
15. O. Reikeras, I. Bjerkreim, A. Kolbenstvedt, Anteversion of the acetabulum and femoral neck in normals and in patients with osteoarthritis of the hip, *Acta Orthop Scand*, 54 (1983) 18–23.
16. R. Tomczak, K. Gunther, T. Pfeifer, H.J. Haberer, A. Rieber, B. Danz, N. Rilling, J.M. Friedrich, H.J. Brambs, [The measurement of the femoral torsion angle in children by NMR tomography compared to CT and ultrasound], *Rofo*, 162 (1995) 224–228.
17. R.J. Hernandez, M.O. Tachdjian, A.K. Poznanski, L.S. Dias, CT determination of femoral torsion, *AJR Am J Roentgenol*, 137 (1981) 97–101.
18. N. Moftian, T. Samad Soltani, Z. Salazadeh, H. Hossein Pour-feizi, Y. Gheibi, A. Fazlollahi, P. Rezaei-Hachesu, Computer-aided Measurement System Using Image Processing for Measuring Cobb Angle in Scoliosis, *Middle East Journal of Rehabilitation and Health Studies*, 9 (2021).
19. J. Schock, D. Truhn, D. Nurnberger, S. Conrad, M.S. Huppertz, S. Keil, C. Kuhl, D. Merhof, S. Nebelung, Artificial intelligence-based automatic assessment of lower limb torsion on MRI, *Sci Rep*, 11 (2021) 23244.
20. N. Gavira, B. Cochard, N. Guanziroli, G. Di Laura Frattura, R. Dayer, D. Ceroni, A new method for assessing tibial torsion using computerized tomography in a pediatric population, *Front Pediatr*, 12 (2024) 1368820.
21. A. Suri, S. Tang, D. Kargilis, E. Taratuta, B.J. Kneeland, G. Choi, A. Agarwal, N. Anabaraonye, W. Xu, J.B. Parente, A. Terry, A. Kalluri, K. Song, C.S. Rajapakse, Conquering the Cobb Angle: A Deep Learning Algorithm for Automated, Hardware-Invariant Measurement of Cobb Angle on Radiographs in Patients with Scoliosis, *Radiol Artif Intell*, 5 (2023) e220158.
22. J.T. Schwartz, B.H. Cho, P. Tang, J. Schefflein, V. Arvind, J.S. Kim, A.H. Doshi, S.K. Cho, Deep Learning Automates Measurement of Spinopelvic Parameters on Lateral Lumbar Radiographs, *Spine (Phila Pa 1976)*, 46 (2021) E671–E678.
23. M.M. Shariatnia, T. Ramazanian, J. Sanchez-Sotelo, H. Maradit Kremers, Deep learning model for measurement of shoulder critical angle and acromion index on shoulder radiographs, *JSES Rev Rep Tech*, 2 (2022) 297–301.
24. Q. Zheng, S. Shellikeri, H. Huang, M. Hwang, R.W. Sze, Deep Learning Measurement of Leg Length Discrepancy in Children Based on Radiographs, *Radiology*, 296 (2020) 152–158.
25. G.P. Lee, S.H. Park, Y.J. Kim, J.W. Chung, K.G. Kim, Enhancing Disease Classification in Abdominal CT Scans through RGB Superposition Methods and 2D Convolutional Neural Networks: A Study of Appendicitis and Diverticulitis, *Comput Math Methods Med*, 2023 (2023) 7714483.
26. A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, 3D Slicer as an image computing platform for the Quantitative Imaging Network, *Magn Reson Imaging*, 30 (2012) 1323–1341.
27. E. Liodakis, I. Doxastaki, K. Chu, C. Krettek, R. Gaulke, M. Citak, M. Kenaway, Reliability of the assessment of lower limb torsion using computed tomography: analysis of five different techniques, *Skeletal Radiol*, 41 (2012) 305–311.
28. S. Waldt, Wortler, K., <Measurements_and_Classifications_in_Musculoskeletal_Lower_Limb_Alignment.pdf>, *Measurements and Classifications in Musculoskeletal Radiology* 2013.

29. N. Pascual-Leone, A. Jahandar, R. Davie, J.T. Bram, D.E. Chipman, C.W. Imhauser, D.W. Green, Femorotibial rotation is linearly associated with tibial tubercle-trochlear groove distance: A cadaveric study, *J ISAKOS*, 9 (2024) 598-602.
30. B.G. Hill, J.D. Krogue, D.S. Jevsevar, P.L. Schilling, Deep Learning and Imaging for the Orthopaedic Surgeon: How Machines "Read" Radiographs, *J Bone Joint Surg Am*, 104 (2022) 1675-1686.
31. N.A. Segal, J.M. Nilges, W.M. Oo, Sex differences in osteoarthritis prevalence, pain perception, physical function and therapeutics, *Osteoarthritis Cartilage*, 32 (2024) 1045-1053.
32. M.E. Rayed, S.M.S. Islam, S.I. Niha, J.R. Jim, M.M. Kabir, M.F. Mridha, Deep learning for medical image segmentation: State-of-the-art advancements and challenges, *Informatics in Medicine Unlocked*, 47 (2024).
33. C. Ries, C.K. Boese, N. Ott, J. Doerner, L.P. Muller, M. Hackl, Intra-individual variance of bilateral femoro-tibial leg rotation: a CT study of 105 healthy subjects, *Knee Surg Sports Traumatol Arthrosc*, 29 (2021) 1106-1113.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

For Approval