RESEARCH ARTICLE

Computed tomography-based automated 3D measurement of femoral version: Validation against standard 2D measurements in symptomatic patients

Florian Schmaranzer[1] | Mohammadreza Movahhedi[2] | Mallika Singh[2] |
Jennifer R. Kallini[2] | Andreas K. Nanavati[3] | Simon D. Steppacher[3] |
Alexander F. Heimann[4] ⬥ | Ata M. Kiapour[2] | Eduardo N. Novais[2] ⬥

[1]Department of Diagnostic, Interventional and Pediatric Radiology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[2]Department of Orthopaedic Surgery and Sports Medicine, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[3]Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[4]Department of Orthopaedic Surgery and Traumatology, HFR – Cantonal Hospital, University of Fribourg, Fribourg, Switzerland

Correspondence
Florian Schmaranzer, Department of Diagnostic, Interventional and Pediatric Radiology, Inselspital, Bern University Hospital, University of Bern, Freiburgstrasse, 3010 Bern, Switzerland.
Email: fschmaranzer@gmail.com

Abstract

To validate 3D methods for femoral version measurement, we asked: (1) Can a fully automated segmentation of the entire femur and 3D measurement of femoral version using a neck based method and a head-shaft based method be performed? (2) How do automatic 3D-based computed tomography (CT) measurements of femoral version compare to the most commonly used 2D-based measurements utilizing four different landmarks? Retrospective study (May 2017 to June 2018) evaluating 45 symptomatic patients (57 hips, mean age 18.7 ± 5.1 years) undergoing pelvic and femoral CT. Femoral version was assessed using four previously described methods (Lee, Reikeras, Tomczak, and Murphy). Fully-automated segmentation yielded 3D femur models used to measure femoral version via femoral neck- and head-shaft approaches. Mean femoral version with 95% confidence intervals, and intraclass correlation coefficients were calculated, and Bland-Altman analysis was performed. Automatic 3D segmentation was highly accurate, with mean dice coefficients of 0.98 ± 0.03 and 0.97 ± 0.02 for femur/pelvis, respectively. Mean difference between 3D head-shaft- ($27.4 \pm 16.6°$) and 3D neck methods ($12.9 \pm 13.7°$) was $14.5 \pm 10.7°$ ($p < 0.001$). The 3D neck method was closer to the proximal Lee ($-2.4 \pm 5.9°$, -4.4 to $0.5°$, $p = 0.009$) and Reikeras ($2 \pm 5.6°$, 95% CI: 0.2 to 3.8°, $p = 0.03$) methods. The 3D head-shaft method was closer to the distal Tomczak ($-1.3 \pm 7.5°$, 95% CI: -3.8 to 1.1°, $p = 0.57$) and Murphy ($1.5 \pm 5.4°$, -0.3 to 3.3°, $p = 0.12$) methods. Automatic 3D neck-based-/head-shaft methods yielded femoral version angles comparable to the proximal/distal 2D-based methods, when applying fully-automated segmentations.

KEYWORDS
deep learning, femoral osteotomy, femoral version, hip arthroscopy

1 | INTRODUCTION

Femoroacetabular impingement syndrome (FAIS) is an established cause of hip pain in young adults, and has been described as a motion-related clinical disorder of the hip caused by premature contact between the proximal femur and acetabulum.[1,2] Cam and pincer deformities have typically been linked with FAIS. However, more recent evidence suggests that rotational abnormalities of the femur can aggrandize/compensate an impingement-conflict or instability problem.[3] Meanwhile, femoral version is now being considered as a third pillar aside of cam and pincer deformities.[4–6] In fact, abnormal femoral version, has been described in 1 of 6 patients who present with hip pain attributed to FAI or hip dysplasia.[6] High femoral version (anteversion) can lead to posterior extra-articular ischiofemoral impingement and in-toeing gait, while low femoral version (retroversion) can lead to anterior intra-articular and extra-articular impingement and out-toeing gait.[7,8] While the effect of abnormal femoral version on outcome following surgical FAI correction and the role of femoral osteotomies is still debated,[9,10] several studies report worse outcomes following hip arthroscopy for those with femoral retroversion or excessive anteversion and borderline hip dysplasia (DDH).[11,12] Besides the hip, femoral anteversion is a major predictor of patellar dislocation and instability in the knee,[13] contributing to the rotational alignment leading to anterior knee pain and can negatively affect postoperative results after total knee arthroplasty.[14]

It is thus imperative that femoral version is considered in the initial diagnostic workup for patients evaluated for joint-preserving hip surgery and certain knee pathologies. Numerous methods for measurement of femoral version have been described with landmarks spanning proximally from the greater trochanter to the lesser trochanter distally.[15–18] This is important as the choice of the measurement method will affect the resulting femoral angles and reference values introducing risk of misdiagnosis if not applied consistently.[19,20] In clinical routine, measurement of femoral version is typically performed on axial images sections through the pelvis and distal femoral condyles. This is problematic as altered/non-standardized patient positioning with changes in hip rotation and flexion can considerably affect the resulting femoral version angles if standard 2D measurement on axial slices are performed.[21–23] Therefore, efforts have been made to minimize these effects related to patient positioning by using 3D models of the entire femur.[24–27] Ideally such measurements would be performed in an automatic fashion, including different anatomical landmarks to reflect the use of the various femoral version measurements and improve consistency across institutions.

Thus, we asked: (1) Can a fully automated segmentation of the entire femur and 3D measurement of femoral version using a neck based method and a head-shaft based method be performed? (2) How do automatic 3D-based CT measurements of femoral version compare to the most commonly used 2D-based measurements utilizing four different landmarks?

2 | METHODS

2.1 | Patients

Following institutional review board approval, a retrospective study was performed in compliance with Health Insurance Portability and Accountability Act guidelines at a tertiary pediatric hospital with a referral center for hip disease (Level of evidence III). Then, a review of the imaging database for a consecutive series of patients who presented at the outpatient clinic at Boston Children's Hospital and had radiographic imaging and pelvic CT between May 2017 and June 2018 was performed. Patients were excluded if they had CT without full coverage of the femur. Patients were not excluded due to surgical hardware or a history of previous hip surgery. Clinical notes were reviewed to confirm the diagnosis of hip pain which was made by a senior pediatric hip surgeon. The diagnosis of hip pain in all patients was based on symptoms persisting for more than 3 months, coupled with a positive impingement test (FADIR) and/or a positive apprehension and the presence of osseous deformities.[28] Accordingly, in a given patient diagnosis of hip pain was either unilateral or bilateral.

Thirty-six patients with external examinations without full coverage of the femur were excluded from the initial sample of 81 patients. Our final cohort was comprised of 45 patients (mean ± standard deviation, 18.7 years ± 5.1; age range, 13–38 years; 18 males) and 57 symptomatic hips (33 patients with unilateral hip pain and 12 patients with bilateral hip pain). FAIS was the most common hip condition present (32 hips [56%]), followed by DDH (17 hips [30%]). Regarding previous surgery, hip arthroscopy was the most common performed procedure (13 hips [54%]), followed by femoral osteotomy (4 hips [17%]). The most common subsequent surgery was periacetabular osteotomy (12 hips [41%]), followed by hip arthroscopy (9 hips [31%]), and surgical hip dislocation (5 hips [17%]) (Table 1).

2.2 | Computed tomography

A dual-source CT scanner (Somatom Force; Siemens Healthcare) or a 64-slice multi–detector row CT scanner (Sensation; Siemens Healthcare) were used for a noncontrast helical CT of the femora and pelvis. Before imaging patients were positioned supinely with knees straight and toes taped together. The field of view included the acetabular roof to the tibial plateau. Dual-source CT scan parameters were the following: collimation, 128 × 0.6 mm; pitch, 2.85; and rotation time, 250 msec. Automated attenuation-based tube current modulation (40-mA reference) for dose reduction, and an automatic kilovoltage selection algorithm was implemented (120-kV reference; dose optimization level, seven). The multi-detector row CT scanner had the following parameters: collimation, 64 × 0.75 mm; pitch, 1.3; and rotation time, 500 msec. Regarding dose reduction, automated attenuation-based tube current modulation (CARE Dose; Siemens Healthcare) was applied (36-mA reference, 120 kVp).

TABLE 1 Demographic characteristics of the study group.

Characteristic	57 Hips (45 patients)	
	Freq.	(%)
Age (*years; mean ± SD*)	18.7	±5.1
Sex (*% female*)	27	(60%)
Hip condition		
Femoroacetabular impingement	32	(56%)
Developmental dysplasia of the hip	17	(30%)
Valgus deformity (neck-shaft angle >139°)	6	(11%)
Sequalae of slipped capital femoral epiphysis	4	(7%)
Perthes-like deformity	4	(7%)
Previous surgery, 19 patients	24	(42%)
Hip arthroscopy	13	(54%)
Femoral osteotomy	4	(17%)
Pelvic osteotomy	2	(8%)
In-situ fixation	2	(8%)
Pelvic and femoral osteotomy	2	(8%)
Surgical hip dislocation	1	(4%)
Subsequent surgery, 26 patients	29	(51%)
Periacetabular osteotomy	12	(41%)
Hip arthroscopy	9	(31%)
Surgical hip dislocation	5	(17%)
Femoral osteotomy	1	(3%)
Surgical hip dislocation and femoral osteotomy	1	(3%)
Surgical hip dislocation and periacetabular osteotomy	1	(3%)

2.3 | 2D measurement of femoral version

A radiologist with 6 years of experience in diagnostic imaging of the hip (F.S.) and a medical student trained on a sample of 20 cases not included in this analysis (J.R.K.) performed the 2D measurement of femoral version on a PC by using the electronic calipers from our picture archiving computer system. The readers were then blinded to the original reports and performed measurements independently on the CT scans, with the radiologist (F.S.) repeating the reading session 4 weeks after initial review. Measurements of both readers were performed independently and without disclosure of the used image slices. Four of the most common measurement methods for femoral version, which are all based on true axial images, were included (Figure 1). Using the femoral head center as first reference these measurements differ regarding the second proximal femoral landmark to define its reference line. The methods implemented starting with the most proximal landmark: Lee et al.,[15] which uses the femoral head

center on the same slice as the proximal femoral neck axis at the level of the grater trochanter; Reikeras et al.[16] with the proximal reference at the level of the femoral neck where the anterior and posterior cortices appear parallel; Tomczak et al.[17] using the center of the greater trochanter at the base of the femoral neck; and Murphy et al.[18] with the proximal femoral axis defined at the base of the femoral neck just above the lesser trochanter (Table 2). For the distal reference, a line connecting the lateral and medial posterior condyles was drawn.

2.4 | 3D automatic measurement of femoral version

Three-dimensional measurement of femoral version was based on an automated two-step approach: First, deep learning CT-based segmentation of the entire femur was done using a supervised deep learning approach. This was followed by automated landmark detection and measurement using in-house developed software. For automatic segmentation UNet based convolutional neural networks (CNN) were developed, trained and validated on manually segment clinical CT scans (*n* = 424; 80% training, 20% testing) with varying degrees of bone deformities (e.g., FAI, developmental dysplasia of the hip, slipped capital femoral epiphysis, Perthes disease and cerebral palsy) from a wide range of males and females at different stages of skeletal maturity (age: 17 ± 9 years, range: 2–61years; 48% females; 22% without hip pathology). The CT images were randomly selected from our institutional radiology database including images obtained between 2000 and 2022. The automatic segmentation pipeline had two independent modules: (1) to segment the femur and (2) to separate femoral neck, femoral head, and femoral epicondyles. The CNNs were then supplemented by a custom-developed postprocessing step to refine the segmentation by removing random components. The final models were exported as fine surface mesh (i.e., each model is defined by a set of faces and vertices) to be used for anatomy measurements.

A custom-developed validated program (VirtualHip^TM, Boston Children's Hospital) was then used to measure femoral version from the reconstructed 3D models. The program used the following rule-based algorithms to automatically define the coordinate systems and to measure femoral version.

(1) Coordinate system definition: Femoral version measurements were conducted in a coordinate system. To establish this, the femoral head center was defined as the center of the best fitted sphere to the femoral head. The best fitted cylinder to the epicondyle model was then used to define medial and lateral epicondyle reference points. The x-axis was defined as the normal vector of the plane passing through the femoral head center and the epicondyle reference points, with the line connecting the femoral head center to the center of the epicondyles as the y axis. The z axis was defined as the vector perpendicular to the x and y axes and would be calculated based on the cross-product between the two (Figure 2).
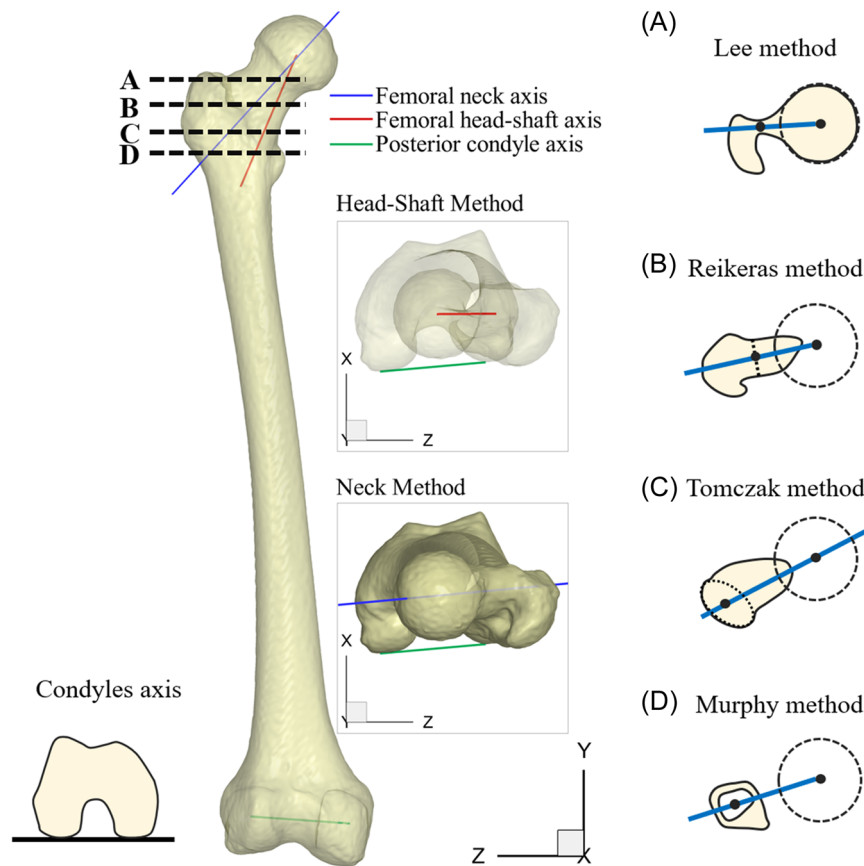
FIGURE 1 Automated 3D measurement of femoral version (neck method and head-shaft method) and the four different 2D measurements of femoral version are shown. For the femoral neck 3D method, the femoral neck axis was defined as the best fitting line between the medial and lateral femoral neck surfaces. For the head-shaft 3D method the angle between the line connecting the femoral head center with the center of the femoral shaft at the level of the lesser trochanter was chosen. Using the femoral head center as the first proximal reference the 2D methods differed regarding the selection of the second proximal reference while the femoral condyle axis served as reference distal reference for all measurement methods alike. (A) For the Lee et al method,[15] the femoral head center is connected on the same slice as the proximal femoral neck axis at the level of the greater trochanter. (B) For the Reikeras et al. method[16] the proximal reference at the level of the femoral neck is selected where the anterior and posterior cortices appear parallel. (C) For the Tomczak et al.17 method the center of the greater trochanter as the base of the femoral neck is chosen. (D) For the Murphy et al.[18] method the proximal femoral axis was defined at the base of the femoral neck just above the lesser trochanter. (A–D) Used with permission from Schmaranzer F, Lerch TD, Siebenrock KA, et al. 2019. Differences in Femoral Torsion Among Various Measurement Methods Increase in Hips With Excessive Femoral Torsion. Clin Orthop Relat Res 477(5), 1073–1083. https://doi.org/10.1097/CORR.0000000000000610.

TABLE 2 Definition of anatomic landmarks for the four measurement methods of femoral torsion.

Methods	2D/3D	Anatomic landmark for definition of the proximal reference axis
Lee method	2D	A line is drawn on the first image on which the femoral head center and the femoral neck can be connected with the most cephalic junction of the greater trochanter
Reikeras method	2D	The femoral head center is connected on superimposed images with the femoral neck axis at the level where the anterior and posterior cortices appear parallel
Tomczak method	2D	The femoral head center is connected on superimposed images with the center of the greater trochanter at the level of the base of the femoral neck
Murphy method	2D	The femoral head center is connected on superimposed images with the base of the femoral neck directly superior to the center of the lesser trochanter
Neck method	3D	Femoral neck axis was defined as the best fitting line between the medial and lateral femoral neck surfaces.
Head-shaft method	3D	Angle between the line connecting the femoral head center with the center of the femoral shaft at the level of the lesser trochanter.
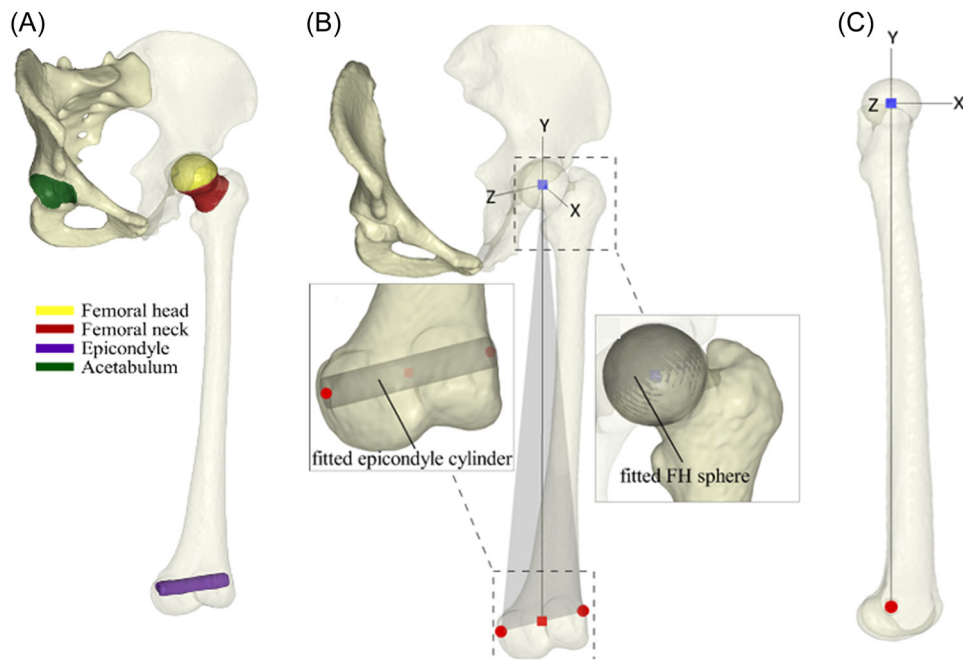
FIGURE 2 (A) Reconstructed 3D models and detected anatomical landmarks. (B) The normal vector of the plane (gray triangle) passes through epicondyle reference points (red circles) and femoral head center (blue square) was used to define the X axis. Magnified view of the epicondyle model, the fitted cylinder, and its reference points. Magnified view of the femoral head, the fitted sphere, and center of the fitted sphere. (C) The line connecting epicondyles center (red square) and femoral head center (blue square) defines the Y axis.

(2) Femoral shaft axis: An iterative process was used to define the femoral shaft axis. First, this was defined using the principal axis of the region of the femur below the neck and above the condyle region (Green Zone, Figure 3A). Planes with a normal vector of the principal axis and with 5 mm intervals starting from the bottom of the region were defined along the axis. The intersections of those planes with the femoral shaft were to establish best fitted circles (Figure 3B). The change in the radius of those fitted circles were used to identify the intertrochanteric region. We then used linear regression to define the femoral shaft as the best fit line to the center of all the fitted circles distal to the lesser trochanter (Figure 3C).

(3) Posterior condyle axis: The posterior condyle axis is defined based on points on the medial and lateral condyles. To find them, the condyle region is first defined by cutting the femur model below the shaft region, followed by local division of the condyle region into two parts in the z direction along the medial-lateral. The most posterior point of each part is selected, and the line which connects these two points was defined as the posterior condyle axis (Figure 4).

(4) Femoral neck axis: The segmented femoral neck model was used to identify the medial and lateral surface of the femoral neck. Similar to the approach used for defining the femoral head shaft, we used an iterative approach to define the femoral neck axis. The femoral neck was first defined as the line connecting the center of the medial and lateral neck surfaces. We then established several intersection planes perpendicular to the femoral neck axis, with the intersection with the femoral neck used to establish best-fit circles. The femoral neck axis was defined as the best-fit line passing through the center of the established circles using linear regression (Figure 5).

(5) Femoral version measurement: Femoral version was measured as the angle between the femoral neck axis and posterior condyle axis in the X-Z (axial) plane (3D neck method). To further replicate the femoral version measurements using more distal landmarks, we also measured femoral version as the angle between the line connecting the femoral head to the center of the femoral shaft at the lesser trochanter level (red circle in Figure 3C) and the posterior axis of the femoral condyles in the axial plane (3D head-shaft method).

2.5 | Statistical analysis

Kolmogorov-Smirnov tests were performed to confirm normal distribution of continuous data.

To answer question one, dice coefficients were calculated as the spatial overlap in percent between the manually- and automatically segmented 3D models, to evaluate the performance of the automatic 3D segmentations.[29,30]

To answer question two, 2D measurements of femoral version were performed twice by one rater and once by a second rater to assess inter- and intra-rater reliability using intra-class correlation coefficients (ICC). In addition, mean intra-observer bias including corresponding standard deviations were calculated for the 2D femoral version measurement methods. For further comparison of the four 2D femoral version measurements mean values from the two readings of reader 1 of each were used. Femoral version angles of the 3D femoral neck method and of the 3D head-shaft method were compared with paired t-tests. Comparison of femoral version angles among the 3D
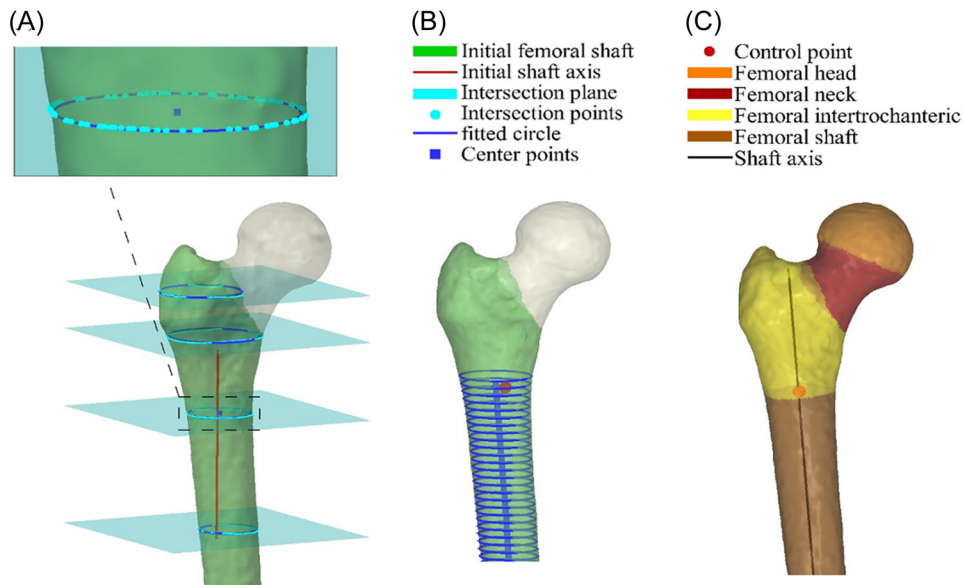
FIGURE 3 Determining the shaft axis and the intertrochanteric region. (A) The zone on the femur that does not belong to the femoral head or neck is initially considered as the femoral shaft. Planes (cyan) with normal vector of the initial shaft axis (obtained from principal axes) with 5 mm intervals starting from the bottom of the region are defined along the axis. (B) The intersections of these planes with femur was used to establish best-fit circles (blue). (C) The red circle (control point) indicates the center of the femoral shaft right below the lesser trochanter. The solid black line indicates the final axis of the femoral shaft.
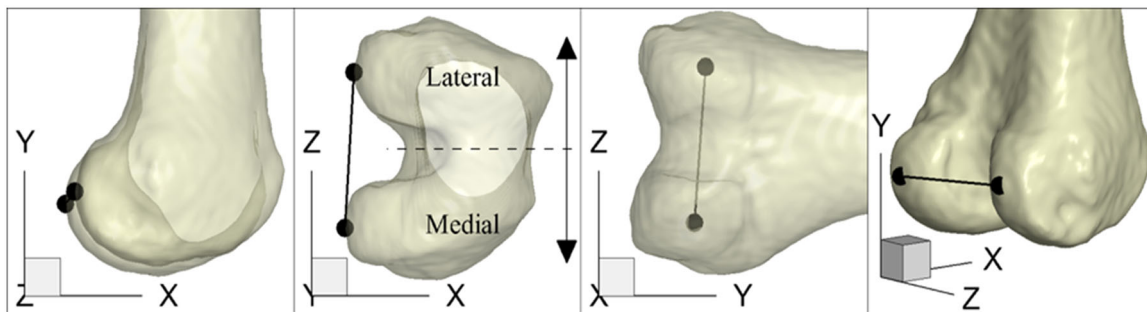
FIGURE 4 Determining the posterior condyle axis. The condyle is locally devided into medial and lateral zones. For each zone, the most posterior point (the point with minimum X) is selected. The line connecting these two points is the posterior condyle axis.
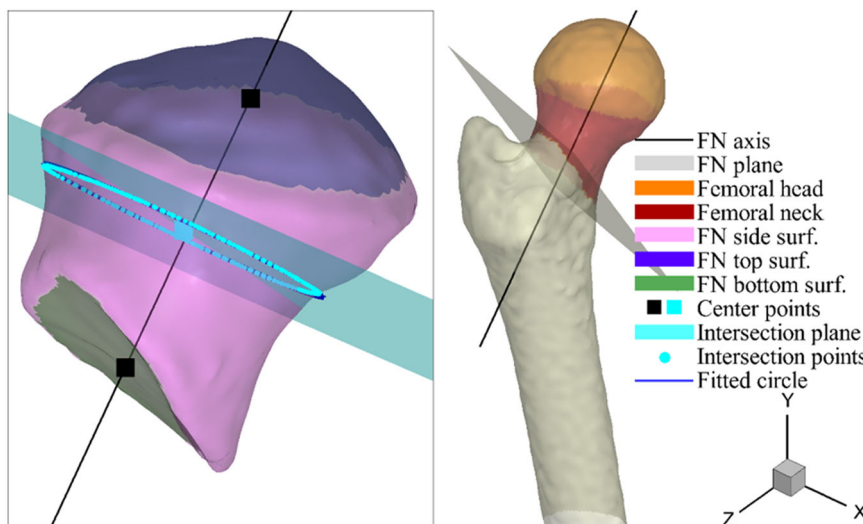
FIGURE 5 Determining the femoral neck (FN) axis. The initial neck axis was established based on the center points of the medial and lateral FN surfaces and was then refined based on the center of the best-fit circles along the neck. Only one of the best-fit circles are shown.

and 2D methods was assessed with MANOVA tests with Bonferroni correction for multiple comparisons. Bland Altman analysis with calculation of mean bias including 95% confidence intervals (CI) and 95% limits of agreement was performed to assess differences between the 3D neck method with the two proximal 2D methods and between the 3D head-shaft method with the two distal 2D methods, respectively. Then the standard deviations (SD) of the intra-rater bias were used as upper and lower limits of the Bland Altman plots to determine the proportion of observed differences which were within the range of intra-observer variation. This was based on the rationale that if most differences between the 2D- and 3D measurement methods were within the range of measurement error of ± 1 SD of the 2D methods they can be considered clinically irrelevant. Statistical analysis was performed using GraphPad Prism (Version 9.1, GraphPad).

3 | RESULTS

3.1 | 3D automatic femoral version measurements

The dice coefficients for automatic segmentation and landmark detection were 0.99 (femur), 0.98 (femoral head), 0.97 (femoral neck) and 0.96 (femoral epicondyles). The mean femoral version for the 3D head-shaft method was higher ($p < 0.001$) with $27.4 \pm 16.6°$ compared to the 3D neck method which yielded a femoral version of $12.9 \pm 13.7°$ (Figure 6).

3.2 | 3D automatic versus 2D CT femoral version methods

The mean 2D femoral version values ranged from $10.5 \pm 14.1°$ for the most proximal Lee method to $28.9 \pm 14.5°$ for the most distal Murphy method (Figure 6).
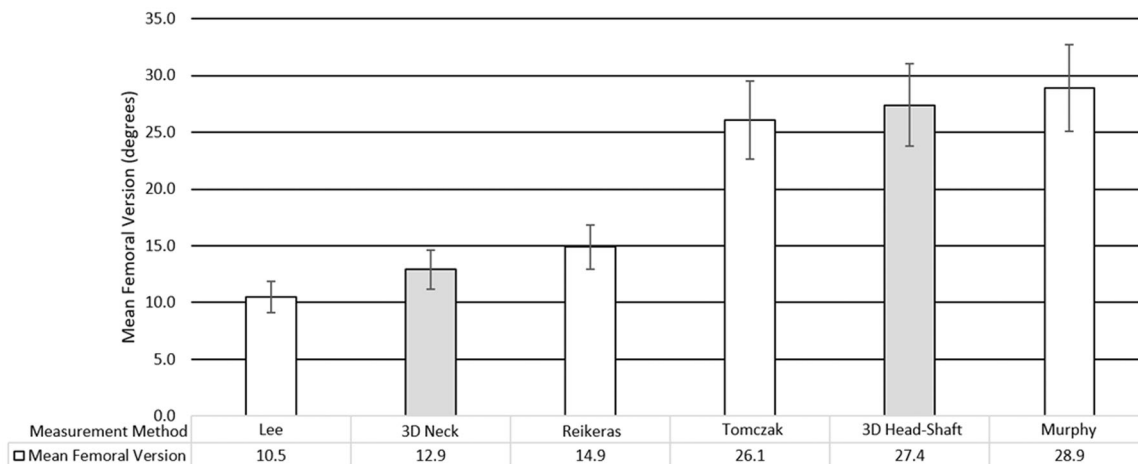
For the 3D neck method, the mean and absolute differences ranged from $2 \pm 5.6°$ (95% CI of 0.2 to 3.8°, $p = 0.03$) and $3.7 \pm 4.6°$ (2.4–4.9°) for the Reikeras method to $16 \pm 7.1°$ (13.7–18.3°, $p < 0.001$)

and $16.3 \pm 6.5°$ (14.6–18.0°) for the most distal method according to Murphy (Table 3). Regarding Bland Altman analysis, 35 of 57 measurement differences (61.4%) between the 3D neck- and the Lee method were within ± 1 SD of the intra-observer variation of 4.3° of the Lee method (Figure 7A). For comparison between the 3D neck- and Reikeras method, this was the case for 39 of 57 (68.4%) measurements which were within ±1 SD of the intra-observer variation of 3.6° of the Reikeras method (Figure 7B).

For the 3D head-shaft method, the mean femoral version angles did not differ when being compared to the distal methods according to Tomczak (−1.3 ± 7.5°, 95% CI of −3.8 to 1.1°, $p = 0.57$) or Murphy (1.5 ± 5.4°, −0.3 to 3.3°, $p = 0.12$). Mean absolute difference was $5.3 \pm 5.4°$ (3.9–6.8°) for the Tomczak method and $4.0 \pm 4.0°$ (2.9–5.0°) for the Murphy method. By contrast, it yielded mean and absolute differences as high as −16.9 ± 7.8° (−19.5 to 14.4°, $p < 0.001$) and 17.1 ± 7.5° (15.1–19.1°) when being compared to the most proximal method according to Lee (Table 3). Regarding Bland Altman analysis, the majority of differences between the 3D head-shaft- and the Murphy method were within the range of intra-observer variation. Between the head-shaft based and Tomczak method, this was not the case as 26 of 57 (45.6%) of measurements were within ± 1 SD of the intra-observer variation of 3.2° of the Tomczak method (Figure 7C). For the Murphy method, 29 of 57 (50.9%) of measurements were within ± 1 SD of the intra-observer variation of 3.0° (Figure 7D). Corresponding automated 3D femoral version measurements and 2D measurement methods of an individual case are shown (Figure 8).

3.3 | Inter and intrarater reliability

Both interrater and intra-rater reliability were excellent across all four 2D measurements of femoral version (all $p < 0.001$). ICCs for interrater reliability ranged from 0.97 (95% CI of 0.95–0.98) with a mean bias of 1.9 ± 4.8° for the Reikeras method to 0.99 (0.98–0.99) with a mean bias of 2.9 ± 3.3 for the method of Murphy. ICCs for intra-rater reliability ranged from 0.98 (0.96–0.99) with a mean bias of 1.7 ± 4.3° for the Lee

Measurement Method	Lee	3D Neck	Reikeras	Tomczak	3D Head-Shaft	Murphy
☐ Mean Femoral Version	10.5	12.9	14.9	26.1	27.4	28.9

FIGURE 6 Assessment of mean femoral version across the four 2D methods and two 3D-based methods.

TABLE 3 Comparison between 3D neck- and 3D head-shaft methods with 2D measurements of femoral version.

Method	3D neck method (95% confidence interval)			3D head-shaft method (95% confidence interval)		
	Mean absolute difference	Mean difference	p value	Mean absolute difference	Mean difference	p value
Lee method	4.3 (3.1 to 5.5)	−2.4 (−4.4 to −0.5)	0.009	17.1 (15.1 to 19.1)	−16.9 (−19.5 to −14.4)	<0.001
Reikeras method	3.7 (2.4 to 4.9)	2 (0.2 to 3.8)	0.03	13.1 (11.0 to 15.1)	−12.5 (−15.4 to −9.7)	<0.001
Tomczak method	13.4 (12.0 to 14.8)	13.2 (11.2 to 15.1)	<0.001	5.3 (3.9 to 6.8)	−1.3 (−3.8 to 1.1)	0.57
Murphy method	16.3 (14.6 to 18.0)	16.0 (13.7 to 18.3)	<0.001	4.0 (2.9 to 5.0)	1.5 (−0.3 to 3.3)	0.12
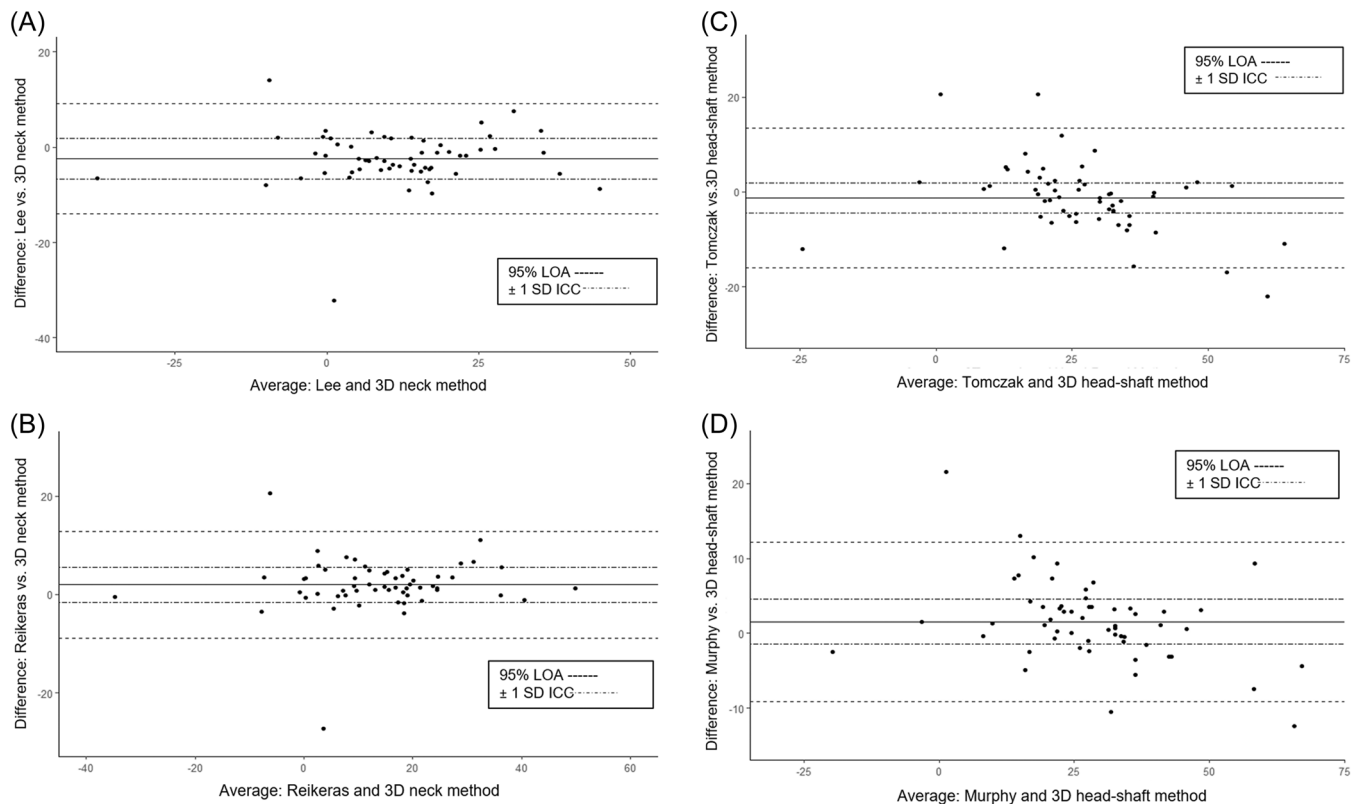
FIGURE 7 (A–D) Bland Altman plots to evaluate the systematic bias between 2D and 3D proximal and distal measurement methods. The solid black line represents the mean difference between the respective 2D and 3D methods. The dark, dashed black line represents the 95% limits of agreement. The lighter dashed line represents 1 standard deviation of the corresponding intra-rater reliability of CT measurements displayed in Table 3. (A) Between the Lee- and 3D neck methods, 35 of 57 (61.4%) measured differences were within ± 1 SD of 4.3°. (B) Between the Reikeras- and 3D neck-methods, 39 of 57 (68.4%) differences were within ± 1 SD of 3.6°. (C) Between the Tomczak- and head-shaft methods, 26 of 57 (45.6%) differences were within ± 1 SD of 3.2°. (D) Between the Murphy- and head-shaft methods, 29 of 57 (50.9%) differences were within ± 1 SD of 3.0°. LOA, limits of agreement.

method to 0.99 (0.98–0.99) with mean biases of 0.4 ± 3.2° and 0.1 ± 3.0° for the methods of Tomczak and Murphy, respectively (Table 4).

4 | DISCUSSION

First, we found our automatic 3D segmentation and landmark detection to be highly accurate, with mean dice coefficients of 0.98 ± 0.01. We observed differences of 14.5 ± 10.7° (p < 0.001) between the automatic 3D neck- and head-shaft methods. Second, when comparing the 3D methods with the 2D methods, we found the

neck method was closer to the more proximal Lee- (−2.4 ± 5.9°, −4.4 to 0.5°, p = 0.009) and Reikeras- (2 ± 5.6°, 95% CI 0.2 to 3.8°, p = 0.03) methods, with the majority of measurements (>50%) being in the range of the measurement error. The head-shaft method showed mean differences closest to the more distal Tomczak (−1.3 ± 7.5°, −3.8 to 1.1°, p = 0.57) and Murphy (1.5 ± 5.4°, −0.3 to 3.3°, p = 0.12) methods. Of these two, the Murphy method yielded differences which were mostly (>50%) within the clinically irrelevant range of intra-observer measurement variation.

Several studies have utilized a segmentation-based 3D approach for measuring femoral version through either semiautomatic[26,31] or
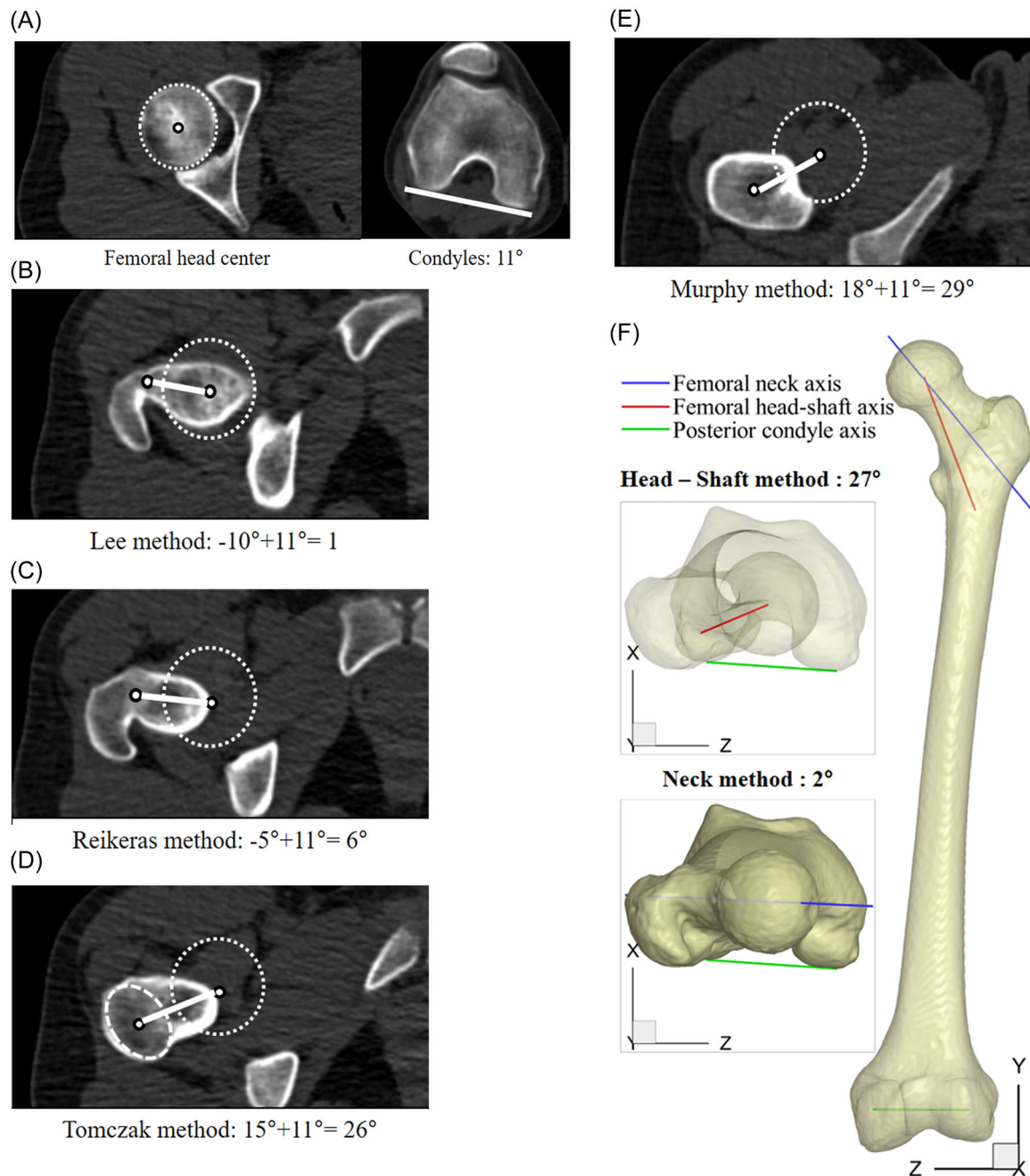
FIGURE 8 A 32-year-old woman undergoing CT for assessment of femoral version using 2D and 3D measurement methods. (A) Section through the femoral rotation center and femoral condyles which serve as the most proximal landmark and the distal reference axis of 11°. (B–E) Femoral version angles increase from the proximal to the distal femoral landmarks: (B) 1° for the Lee method, (C) 6° for the Reikeras method, (D) 26° for the Tomczak method, (E) 29° for the Murphy method. (F) Segmentation of the entire femur for automated 3D measurement of femoral version is shown. The proximal 2D methods of Lee and Reikeras compared well to the 3D femoral neck version of 2° while the distal 2D methods compared well to the 3D head shaft method of 27°.

automatic approaches.[25,27] Schock and colleagues[25] used a deep learning approach applying a convolutional neural network (U-net) to segment the femur and tibia for assessment of femoral version on MRI. Like our study, they reported high segmentation accuracy, with Dice coefficients ranging from 0.89 ± 0.02 to 0.93 ± 0.02 for the proximal and distal landmarks of the entire femur and tibia.[25] Based on these 3D models, mean femoral version was 15.8°, comparable to

our proximal, 3D neck method which yielded 12.9° of mean femoral version. Berryman et al.[26] implemented a semi-automatic method with a neck-fitting approach using point clouds of the neck surface to define the femoral neck axis. Tested on cadaveric femora using CT, this approach gave a femoral version value of 19.1 ± 7.3°, which falls between both our 3D head-shaft and neck methods. Casciaro and Craiem,[27] used a fully automatic approach based on cylinder-fitting

TABLE 4 Inter and intrarater reliability for 2D femoral version methods.

Interrater reliability					
Method	**ICC**	**95%CI**	**MD**	**±SD**	**95%CI**
Lee	0.98	(0.97–0.99)	2.6	±3.6	(−2.8 to 8.0)
Reikeras	0.97	(0.95–0.98)	1.9	±4.8	(−3.4 to 7.2)
Tomczak	0.98	(0.97–0.99)	3.6	±3.4	(−1.6 to 8.8)
Murphy	0.99	(0.98–0.99)	2.9	±3.3	(−2.5 to 8.3)
Intrarater reliability					
Method	**ICC**	**95%CI**	**MD**	**±SD**	**95%CI**
Lee	0.98	(0.96–0.99)	1.7	±4.3	(−3.6 to 7)
Reikeras	0.98	(0.97–0.99)	0.6	±3.6	(−4.6 to 5.8)
Tomczak	0.99	(0.98–0.99)	0.4	±3.2	(−4.7 to 5.6)
Murphy	0.99	(0.98–0.99)	0.1	±3.0	(-5.3 to 5.5)

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; MD, mean difference; SD, standard deviation.

and compared it to 2D manual measurement, in dried human femurs using CT. They reported femoral anteversion of $19.8 ± 6.6°$, comparable to Berryman et al.[26] Finally, Häller et al.[31] employed a semi-automatic method using global thresholding and segmentation of the femur and showed comparable measurements between 3D MRI and CT with mean differences of $0.4 ± 2.8°$ ($p = 0.253$). In contrast to our study, however, they did not segment the entire femur. This can be problematic due to variations in leg positioning during scanning, especially for patients in whom neutral hip positioning may not be possible as a result of pain or rigidity (SCFE patients, patients with neuromuscular disease).[21] Guidetti and colleagues[23] documented this occurrence, reporting decrease in femoral version by 1° per every degree of hip flexion and increasing femoral version by 0.35° per every degree of adduction. When compared to stereoradiography, CT measurements of femoral version in which limb position was altered showed variations as high as −9 to 12.5° for changes in hip position as small as 5° of hip flexion and 10° of hip extension in cadaveric femurs.[22]

To the best of our knowledge, this is the first study to clinically apply a fully automated deep learning approach for comparison of 3D methods based on proximal and distal approaches with the most commonly used 2D measurement methods. Our neck-based method had a mean version ($12.9 ± 13.7°$) that closely approximated the two most proximal 2D methods of Lee ($10.5 ± 14.1°$) and Reikeras ($14.9 ± 14.0°$). This is sensible, given that our 3D femoral neck method uses the center of the femoral neck axis (between the medial and lateral cortices) as the reference point, similar to the approximation of the femoral neck axis when applying the methods described by Reikeras et al.[16] By contrast, 3D head-shaft based femoral version angles ($27.4 ± 16.6°$) were comparable to the distal 2D methods of Tomczak ($26.1 ± 13.8°$) and Murphy ($28.9 ± 14.5°$). This too, is a reflection of the landmarks used. The head-shaft method used the line connecting the center of the femoral head to the center of the

femoral shaft at the lesser trochanter level, which is closest to the method described by Murphy and colleagues.[18] The mean femoral version we reported for the four 2D methods are consistent with previous studies' values, such as Schmaranzer et al.[20] (Lee: $11 ± 11°$, Reikeras: $15 ± 11°$, Tomczak: $25 ± 12°$, and Murphy: $28 ± 13°$), and Berryman et al.[26] (Lee: $8.1 ± 6°$, Reikeras: $12.4 ± 7.4°$, and Murphy: $22.6 ± 8.7°$). Accordingly, our findings further confirm that femoral version angles increase with more distal landmarks being used for 2D methods and 3D methods alike.

Other comparisons between 3D and 2D version methods exist, yet only a single 3D approach was validated in these studies. Based on a more proximal reference, Casciaro and Craiem[27] demonstrated that their automatic 3D method had a mean version ($19.8 ± 6.6°$) closest to the 2D Reikeras method ($19.3 ± 6.4°$). Schock et al.,[25] also only validated a proximal 3D version method against 2D methods and showed mean 3D femoral version of 15.8° closest to the Lee method (reader 1: 16.1°, reader 2: 18.0°). Finally, Berryman and colleagues[26] were the only ones to validate a distal 3D approach, with their semi-automatic neck-fitting method having a mean version ($19.1 ± 7.3°$) which approximated the femoral version angles of the Murphy method ($22.6 ± 8.7°$).

Having consistent methodology for measuring femoral version is imperative, given the previously reported variation of up to 20°[20,32] depending on landmark selection which was further confirmed in our study. Of note, this difference further increases at the extremes of femoral anteversion and valgus deformity.[33] In addition to the four common 2D methods described in our study, there are several other 2D ones such as: Hernandez-,[34] Weiner-,[35] Yoshioka-,[36] Jarrett-,[37] and Waidelich methods.[38] The considerable institutional variability regarding used measurement method is reflected in a recent systematic review by Sinkler and colleagues.[39] The authors included 18 studies to assess the impact of abnormal femoral version on outcomes of arthroscopic hip surgery. Among these studies, 7 did not specifically describe the applied femoral version measurement method. For the remaining 11 studies, five different methods were applied. This variability and inconsistent reporting of the applied measurement methods yields the risk of misdiagnosis and confusion on the parts of medical specialists involved in diagnosis and treatment and makes validation and comparison between studies challenging. Thus, an automatic 3D measurement approach like those we have developed may enable an objective and standardized measurement of femoral version, while still accounting for the distinction between commonly used proximal and distal references. Such a combined automated approach would further enable us to systemically assess the location of the rotational deformity of the proximal femur in different hip deformities such as post-slip deformities and in hips with coxa valga et antetorta.

This study has a number of limitations. First, we did not have a ground truth to directly compare with, such as cadaveric femora. However, we used an actual young patient cohort whereas the majority of proposed 3D femoral version measurements were solely based on cadaveric specimen.[22,26,27] In addition, we used the standard deviation of the intra-observer variation of four commonly

used measurement methods as a measure of clinical relevance bias between manual 2D- and automated 3D methods. Also, we did not perform an additional validation step comparing between manual 2D-, automated 3D measurements against manual 3D-based measurements of femoral version. This would have required additional software interfaces for manual annotations of the 3D models. Given the high segmentation accuracy of the proximal femur (dice coefficients 0.97–0.98) we consider this potential source of bias negligible. Second, our study used only the four most common 2D methods which reflect the most proximal and distal landmarks described for femoral version analysis based on true axial images. Despite being frequently used, we chose not to include the Jarrett method,[37] which uses the femoral neck axis on axial oblique slices as a reference and thereby introduces a systematic bias through the oblique plane of image acquisition.[40] Third, we cannot recommend one femoral version measurement over another based on the presented data. This is related to the fact that to date no outcome-based data supporting use of either method is available. However, there is evidence suggesting distal measurement methods yield a more accurate assessment of femoral version in patients with coxa valga and antetorta.[33] By providing and validating two 3D based measurement methods, we sought to cut down the number of available femoral version methods currently used and simplify it with a fully automated approach. Finally, we could not actually confirm that our automated 3D methods for measuring femoral version are independent from hip position since image acquisition was standardized with knees straight and toes taped together. This prevented alterations in hip position which could have affected resulting version angles and inter- and intra-rater reliability. In summary, we found that our automatic deep learning-based 3D neck- and head-shaft methods yielded femoral version angles comparable to the proximal and distal 2D-based methods, respectively, validating them for clinical use. Application of automated 3D femoral version measurements with the presented approach can pave the way towards a more standardized, simplified, and objective analysis of rotational malalignment of the femur in the diagnostic workup ahead of joint preserving hip surgery independent of measurement error. Use of the proposed method may in turn help to identify the measurement method which best reflects hip function and should be used for surgical planning.

AUTHOR CONTRIBUTIONS

Florian Schmaranzer: Significant contributions to the research design or to the acquisition, analysis, or interpretation of data and drafting of the paper or its critical revision. **Mohammadreza Movahhedi**: Substantial contributions to the research design or to the acquisition, analysis, or interpretation of data. **Mallika Singh**: Substantial contributions to the research design or to the acquisition, analysis, or interpretation of data. **Jennifer R. Kallini**: Substantial contributions to the research design or to the acquisition and analysis of data. **Andreas K. Nanavati**: Draft of the paper or its critical revision. **Simon D. Steppacher**: Substantial contributions to the research design or to the acquisition, analysis, or interpretation of data. **Alexander F. Heimann**: Draft of the paper or its critical revision. **Ata M. Kiapour**: Significant contributions to the research design or to the acquisition, analysis, or interpretation of data and drafting of the paper or its critical revision. **Eduardo N. Novais**: Draft of the paper or its critical revision. Each author certifies that he has read and approved the final submitted manuscript.

ACKNOWLEDGMENTS

ORCID

Alexander F. Heimann http://orcid.org/0000-0001-6578-5759
Eduardo N. Novais http://orcid.org/0000-0002-9187-3100

REFERENCES

1. Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res*. 2003;417(417):112-120.
2. Griffin DR, Dickenson EJ, O'Donnell J, et al. The warwick agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement. *Br J Sports Med*. 2016;50(19): 1169-1176.
3. Hanke MS, Schmaranzer F, Steppacher SD, Lerch TD, Siebenrock KA. Hip preservation. *EFORT Open Rev*. 2020;5(10):630-640.
4. Schmaranzer F, Kheterpal AB, Bredella MA. Best practices: hip femoroacetabular impingement. *Am J Roentgenol*. 2021;216(3): 585-598.
5. Novais EN, Movahhedi M, Kiapour AM, Bixby SD. Excessive femoral anteversion leading to symptomatic posterior femoroacetabular impingement, cam deformity of the posterior femoral Head-Neck junction, and anterior hip instability in a dancer: a case report. *JBJS Case Connector*. 2023;13(3):1-7. https://journals.lww.com/10.2106/JBJS.CC.22.00794
6. Lerch TD, Todorski IAS, Steppacher SD, et al. Prevalence of femoral and acetabular version abnormalities in patients with symptomatic hip disease: a controlled study of 538 hips. *Am J Sports Med*. 2018;46(1):122-134.
7. Lerch TD, Eichelberger P, Baur H, et al. Prevalence and diagnostic accuracy of in-toeing and out-toeing of the foot for patients with abnormal femoral torsion and femoroacetabular impingement: implications for hip arthroscopy and femoral derotation osteotomy. *Bone Jt J*. 2019;101-B:1218-1229.
8. Siebenrock KA, Steppacher SD, Haefeli PC, Schwab JM, Tannast M. Valgus hip with high antetorsion causes pain through posterior extraarticular FAI. *Clin Orthop Relat Res*. 2013;471(12): 3774-3780.
9. Lerch TD, Schmaranzer F, Steppacher SD, Ziebarth K, Tannast M, Siebenrock KA. Most of patients with femoral derotation osteotomy for posterior extraarticular hip impingement and high femoral version would do surgery again. *HIP Int*. 2022;32(2):253-264.
10. Buly RL, Sosa BR, Poultsides LA, Caldwell E, Rozbruch SR. Femoral derotation osteotomy in adults for version abnormalities. *J Am Acad Orthop Surg*. 2018;26(19):e416-e425.

11. Fabricant PD, Fields KG, Taylor SA, Magennis E, Bedi A, Kelly BT. The effect of femoral and acetabular version on clinical outcomes after arthroscopic femoroacetabular impingement surgery. *J Bone Jt Surg*. 2015;97(7):537-543.

12. Chaharbakhshi EO, Hartigan DE, Perets I, Domb BG. Is hip arthroscopy effective in patients with combined excessive femoral anteversion and borderline dysplasia? A match-controlled study. *Am J Sports Med*. 2019;47(1):123-130.

13. Qiao Y, Zhang X, Xu J, Xu C, Zhao S, Zhao J. Internal torsion of the knee: an embodiment of lower-extremity malrotation in patients with patellar instability. *J Bone Jt Surg*. 2022;104(13):1179-1187.

14. Rhee SJ, Cho JY, Jeung SH, Poon KB, Choi YY, Suh JT. Combined rotational alignment change after total knee arthroplasty in different tibial component designs: implications for optimal tibial component rotational alignment. *Knee Surg Relat Res*. 2018;30(1):74-83.

15. Lee YS, Oh SH, Seon JK, Song EK, Yoon TR. 3D femoral neck anteversion measurements based on the posterior femoral plane in ORTHODOC system. *Med Biol Eng Comput*. 2006;44(10):895-906.

16. Reikerås O, Bjerkreim I, Kolbenstvedt A. Anteversion of the acetabulum and femoral neck in normals and in patients with osteoarthritis of the hip. *Acta Orthop Scand*. 1983;54(1):18-23.

17. Tomczak RJ, Guenther KP, Rieber A, Mergo P, Ros PR, Brambs HJ. MR imaging measurement of the femoral antetorsional angle as a new technique: comparison with CT in children and adults. *Am J Roentgenol*. 1997;168(3):791-794.

18. Murphy SB, Simon SR, Kijewski PK, Wilkinson RH, Griscom NT. Femoral anteversion. *J Bone Jt Surg*. 1987;69(8):1169-1176.

19. Schmaranzer F, Kallini JR, Miller PE, Kim YJ, Bixby SD, Novais EN. The effect of modality and landmark selection on MRI and CT femoral torsion angles. *Radiology*. 2020;296(2):381-390.

20. Schmaranzer F, Lerch TD, Siebenrock KA, Tannast M, Steppacher SD. Differences in femoral torsion among various measurement methods increase in hips with excessive femoral torsion. *Clin Orthop Relat Res*. 2019;477(5):1073-1083.

21. Iwasaka-Neder J, Bixby SD, Bedoya MA, et al. Virtual 3D femur model to assess femoral version: comparison to the 2D axial slice approach. *Pediatr Radiol*. 2023;53(12):2411-2423.

22. Morvan G, Guerini H, Carré G, Vuillemin V. Femoral torsion: impact of femur position on CT and stereoradiography measurements. *Am J Roentgenol*. 2017;209(2):W93-W99.

23. Guidetti M, Lambers FM, Kohnen M, et al. Influence of limb positioning during image acquisition on femoral torsion measurements: implications for surgical planning. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2023;11(3):718-729.

24. Ratner D, Kolaczko JG, Jeffers K, et al. Three-dimensional analysis versus two-dimensional slice-based analysis of CT for measuring femoral torsion and its correlation to passive hip range of motion. *Cureus*. 2022;14:1-9. https://www.cureus.com/articles/114055-three-dimensional-analysis-versus-two-dimensional-slice-based-analysis-of-ct-for-measuring-femoral-torsion-and-its-correlation-to-passive-hip-range-of-motion

25. Schock J, Truhn D, Nürnberger D, et al. Artificial intelligence-based automatic assessment of lower limb torsion on MRI. *Sci Rep*. 2021;11(1):23244.

26. Berryman F, Pynsent P, McBryde C. A semi-automated method for measuring femoral shape to derive version and its comparison with existing methods. *Int J Numer Method Biomed Eng*. 2014;30(11):1314-1325.

27. Casciaro ME, Craiem D. Towards automatic measurement of anteversion and neck–shaft angles in human femurs using CT images. *Comput Methods Biomech Biomed Engin*. 2014;17(2):128-136.

28. Steppacher SD, Anwander H, Zurmühle CA, Tannast M, Siebenrock KA. Eighty percent of patients with surgical hip dislocation for femoroacetabular impingement have a good clinical result without osteoarthritis progression at 10 years. *Clin Orthop Relat Res*. 2015;473(4):1333-1341.

29. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297-302.

30. Schmaranzer F, Helfenstein R, Zeng G, et al. Automatic MRI-based three-dimensional models of hip cartilage provide improved morphologic and biochemical analysis. *Clin Orthop Relat Res*. 2019;477(5):1036-1052.

31. Häller TV, Schenk P, Jud L, Hoch A, Götschi T, Zingg PO. Consistency of 3D femoral torsion measurement from MRI compared to CT gold standard. *BMC Musculoskelet Disord*. 2021;22(1):739.

32. Scorcelletti M, Reeves ND, Rittweger J, Ireland A. Femoral anteversion: significance and measurement. *J Anat*. 2020;237(5):811-826.

33. Schmaranzer F, Meier MK, Lerch TD, et al. Coxa valga and antetorta increases differences among different femoral version measurements: potential implications for derotational femoral osteotomy planning. *Bone Jt Open*. 2022;3(10):759-766.

34. Hernandez R, Tachdjian M, Poznanski A, Dias L. CT determination of femoral torsion. *Am J Roentgenol*. 1981;137(1):97-101.

35. Weiner DS, Cook AJ, Hoyt WA, Oravec CE. Computed tomography in the measurement of femoral anteversion. *Orthopedics*. 1978;1(4):299-306.

36. Yoshioka Y, Cooke TDV. Femoral anteversion: assessment based on function axes. *J Orthop Res*. 1987;5(1):86-91.

37. Jarrett DY, Oliveira AM, Zou KH, Snyder BD, Kleinman PK. Axial oblique CT to assess femoral anteversion. *Am J Roentgenol*. 2010;194(5):1230-1233.

38. Waidelich HA, Strecker W, Schneider E. Computed tomographic torsion-angle and length measurement of the lower extremity. The methods, normal values and radiation load. *RoFo*. 1992;157(3):245-251.

39. Sinkler MA, Magister SJ, Su CA, Salata MJ. Femoral version May impact hip arthroscopy outcomes in select patient populations: a systematic review. *Arthroscopy*. 2023;39(1):114-127.

40. Sutter R, Dietrich TJ, Zingg PO, Pfirrmann CWA. Assessment of femoral antetorsion with MRI: comparison of oblique measurements to standard transverse measurements. *Am J Roentgenol*. 2015;205(1):130-135.