Automated Morphometric Analysis of the Hip Joint on MRI from the German National Cohort Study

Marc Fischer, MSc • Sven S. Walter, MD • Tobias Hepp, MD • Manuela Zimmer, BSc • Mike Notohamiprodjo, MD • Fritz Schick, MD, DSc • Bin Yang, DEng

From the Institute of Signal Processing and Systems Theory, University of Stuttgart, Pfaffenwaldring 47, 70550 Stuttgart, Germany (M.F., M.Z., B.Y.); Department of Diagnostic and Interventional Radiology, Section on Experimental Radiology, University Hospital Tübingen, Tübingen, Germany (S.S.W., T.H., M.N., F.S.); and Empirical Inference Department, Max Planck Institute for Intelligent Systems, Tübingen, Germany (T.H.). Received September 4, 2020; revision requested October 10; revision received May 3, 2021; accepted May 17. Address correspondence to M.F. (e-mail: marc.fischer@iss.uni-stuttgart.de).

Supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant 325028047).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2021; 3(5):e200213 • <https://doi.org/10.1148/ryai.2021200213> • Content codes:

Purpose: To develop and validate an automated morphometric analysis framework for the quantitative analysis of geometric hip joint parameters in MR images from the German National Cohort (GNC) study.

Materials and Methods: A secondary analysis on 40 participants (mean age, 51 years; age range, 30–67 years; 25 women) from the prospective GNC MRI study (2015–2016) was performed. Based on a proton density–weighted three-dimensional fast spin-echo sequence, a morphometric analysis approach was developed, including deep learning–based landmark localization, bone segmentation of the femora and pelvis, and a shape model for annotation transfer. The centrum-collum-diaphyseal, center-edge (CE), three alpha angles, head-neck offset (HNO), and HNO ratio along with the acetabular depth, inclination, and anteversion were derived. Quantitative validation was provided by comparison with average manual assessments of radiologists in a cross-validation format. Paired-sample *t* tests with a Bonferroni-corrected significance level of .005 were employed alongside mean differences and 10th/90th percentiles, median absolute deviations (MADs), and intraclass correlation coefficients (ICCs).

Results: High agreement in mean Dice similarity coefficients was achieved (average of $97.52\% \pm 0.46$ [standard deviation]). The subsequent morphometric analysis produced results with low mean MAD values, with the highest values of 3.34° (alpha 03:00 o'clock position) and 0.87 mm (HNO) and ICC values ranging between 0.288 (HNO ratio) and 0.858 (CE) compared with manual assessments. These values were in line with interreader agreements, which at most had MAD values of 4.02° (alpha 12:00 o'clock position) and 1.07 mm (HNO) and ICC values ranging between 0.218 (HNO ratio) and 0.777 (CE).

Conclusion: Automatic extraction of geometric hip parameters from MRI is feasible using a morphometric analysis approach with deep learning.

Supplemental material is available for this article.

© RSNA, 2021

With the advent of large epidemiologic cohorts, studies have indicated the prevalence of abnormalities in the hip joint of asymptomatic volunteers (1). Performing a reliable detection and quantification of subtle characteristics enables a nuanced clinical investigation. However, careful manual assessments are necessary for an accurate identification of femoroacetabular impingement and other precursors of osteoarthritis (2). For imaging data, this assessment comes with high expenses in both time and personnel. Furthermore, the measurement of suitable parameters can vary based on the posed view (3), with a common epiphenomenon being inter- and intrareader variations (4,5).

Advances in three-dimensional imaging have made these assessments increasingly popular (6). In recent years, various bone models have been developed to provide computer-aided diagnosis, making the assessment of morphologic parameters more robust with the aid of section- and orientation-independent hip joint or femur representations (7–15). Procedures with varying degree of automation have been proposed (16–18) to provide consistent quantification of specific parameters across individual participants.

For most aforementioned methods, suitable imaging modalities, such as CT, have become prevalent to provide manual or automated delineations. Solutions relying on MRI can be an adequate replacement (14,19) and bring the benefit of linking potential morphometric findings with prevalent tissue characteristics. Naturally, the medical image analysis community is continually working toward creating automated frameworks (20), but past approaches have remained uncommon due to their complexity and limited applicability. Only recent progress in the field of deep learning (21) has made it possible to reliably process volumes with large interparticipant variability of anatomy, texture, and intensity characteristics (22–24).

With large-scale epidemiologic MRI cohort studies, such as the UK Biobank (25) and German National Cohort (GNC) (26) MRI studies, fully automated quantification becomes not just an opportunity, but a necessity. Leveraging recent progress, this study formulated and evaluated an integrated approach to enable quantitative morphometric analyses of the hip joint from an intricate fat-saturated proton density–weighted MRI sequence of the

Abbreviations

APP = anterior pelvic plane, ASSD = average symmetric surface distance, CCD = centrum-collum-diaphyseal, CE = center-edge, GNC = German National Cohort, H = highest fold mean, HD = Hausdorff distance, HNO = head-neck offset, ICC = intraclass correlation coefficient, L = lowest fold mean, MAD = median absolute deviation

Summary

An integrated deep learning-based morphometric analysis approach for the automated derivation of multiple geometric hip joint parameters was shown to be comparable to interreader agreements in three-dimensional cohort data from the German National Cohort MRI study.

Key Points

- Development of a fully automated approach for deriving geometric hip joint parameters from MRI was feasible on data from the German National Cohort.
- Automated image processing with deep learning-based neural networks achieved accurate anatomic landmark localization (mean distance: $1.23 \text{ mm} \pm 0.83$) and bone segmentation (mean Dice similarity coefficient: $97.52\% \pm 0.46$ across femur and pelvis segmentations) on intricate MRI data.
- A comparison of the automated morphometric analysis with manual measurements shows low mean median deviations, with the highest values consisting of 3.34° (on alpha 03:00 o'clock position) and 0.87 mm (on head-neck offset).

Keywords

Computer-Aided Diagnosis (CAD), Interventional-MSK, MR-Imaging, Neural Networks, Skeletal-Appendicular, Hip, Anatomy, Computer Applications-3D, Segmentation, Vision, Application Domain, Quantification

cross-sectional GNC study. The approach comprised (a) deep learning algorithms, suitable for the localization of prominent landmarks and subsequent semantic segmentation of hip bones and (b) shape model processing providing annotation transfer for reference frame and primitive alignment as well as the calculation of relevant hip parameters. A range of common geometric parameters of the hip joint were derived without human intervention at inference time. The approach and its intermediate results were validated by comparison with manual assessments by radiologists.

Materials and Methods

Study Design

A fully automated analysis framework, including image processing, shape model generation, geometric primitive alignment, and hip parameter derivation was designed and evaluated. Algorithmic implementation details can be found in Appendix E1 (supplement). The centrum-collum-diaphyseal (CCD) angle, center-edge (CE) angle, head-neck offset (HNO) and HNO ratio, acetabular depth, inclination, and anteversion, as well as the alpha angle at specific femoral clock positions, were calculated according to definitions in the respective literature and compared with manual assessments. An illustration of the geometric hip parameters is shown in Figure 1.

Study Population

The GNC (26) is a large, multicentric, prospective cohort study that includes 30 000 individuals of the general population examined with whole-body MRI. In this secondary analysis, 80 participants were drawn randomly from data acquired between 2015 and 2016. Exclusion criteria applied to 20 participants: a field of view cutting off in close proximity of the minor trochanter ($n = 9$); a partial coverage of anterior pelvic areas ($n = 7$); and large cystic lesions ($> 1 \text{ cm}$) in the bone tissue ($n = 4$). A total of 60 individuals were included in this study. Approval from an ethics committee and participant consent have been obtained.

A total of 20 individuals were used for algorithmic development and determination of hip parameter derivation routines. The remaining disjoint subset of 40 participants (mean age, 51 years; age range, 30–67 years; 25 women, and mean body mass index, $26.6 \text{ kg/m}^2 \pm 4.3$) was evaluated by cross-validation, resulting in four folds of 30 training and 10 validation individuals each.

Image Acquisition

Fat-saturated proton density-weighted fast spin-echo images, acquired with a 3-T MRI scanner (Magnetom Skyra; Siemens Healthcare), were considered despite the intricate texture and intensity inhomogeneities, partial volume effects, and partial femoral shaft visibility (see Fig 2A, 2B), because the sequence provides the highest available resolution of $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ containing the hip joint area. The image volumes were acquired with a matrix size of $384 \times 264 \times 160$ and imaging parameters of echo time of 33 msec, repetition time of 1200 msec, and a bandwidth of 500 Hz/pixel.

Hip Parameter Measurements

Hip parameter measurements were performed by two independent radiologists (S.S.W., with 4 years of experience and T.H., with 3 years of experience). In addition, manual annotations (ground truth) for learning-based algorithms, shape model generation, and validation of intermediate results (M.F.) were generated. These annotations included voxel-wise segmentation masks, anatomic landmarks, and geometric surrogate primitives for all individuals, as well as a region-based mesh segmentation on MR images from one woman and man in each fold. The mesh segmentations indicate the femoral head, neck, and shaft, as well as a separate acetabular margin region (Fig 2C, 2D). The landmarks depict the anterior pelvic plane (APP) (Fig 2D), which allows establishment of a reference acetabular orientation (10).

Image Processing

To delineate the bones of interest, two convolutional neural networks were employed, which follow the same base architecture (Fig 3A). The first network predicted the positions of predefined anatomic landmarks (Fig 3D), which were in turn provided to the second network (Fig 3C) to predict segmentation masks. The patch-based architecture (in: $68 \times 68 \times 68$, out: $30 \times 30 \times 30$) was analogous to architectures that were proven successful for the segmentation of different MRI sequences (22,27). In this

Figure 1: Automatically derived geometric hip parameters in this study: **(A)** centrum-collum-diaphyseal angle, **(B)** center-edge angle, **(C)** head-neck offset and head-neck offset ratio, **(D)** alpha angles at the 12:00, 01:30, and 03:00 femoral clock position, **(E)** acetabular anteversion, **(F)** acetabular inclination, and **(G)** acetabular depth.

Figure 2: **(A, B)** Varying imaging characteristics and cutoff heights of the imaging data of the German National Cohort of the left femur, **(C)** region annotations used in the geometric parameter derivation of the femur consisting of head, neck, and shaft, and **(D)** region (acetabular margin) and point annotations (anterior pelvic plane [APP] points) of the pelvis.

work, convolutions with residual connections, efficient spatial pyramid blocks (28) providing dilated convolutions, and convolutional block attention modules (29) were combined, leading to attentive spatial pyramid blocks that replace convolutions in the encoding and decoding branches.

To compensate for the patch-based context loss and aid in differentiating tissues of similar characteristics, positional information of both femur head centers was used to provide spatial reference to the segmentation network. The localization network identified these positions by means of deep regression voting (30,31). Voxel-wise offset vectors that point toward a landmark

position were predicted within a spherical extent. The vectors were aggregated into heatmaps with the maximum being the identified landmark position. Both positions together with each patch center were passed to a dynamic filter network (32) applying a dynamic convolution to the bottleneck of the main segmentation network (Fig 3C).

Morphometric Analysis

Shape model.— Surface mesh models were generated from the predictions. An annotated template was employed to transfer

Figure 3: Employed neural networks: **(A)** base architecture consisting of four levels of encoder-decoder branches including attentive spatial pyramid (ASP) blocks with dilation and attention mechanisms, **(B)** localization network with two multichannel output volumes for foreground and offset vectors, and **(C)** segmentation network with dynamic filter network (DFN) branch and one multichannel output segmentation mask volume.

annotations to shape models of the respective validation fold. Therefore, region annotations, as well as APP points, of the training data were incorporated in respective template models. By means of a point-to-point registration algorithm (33), vertices of bone meshes were aligned. The aligned vertices allowed for the establishment of a mean shape, despite varying cutoff heights of the present femurs. The resulting mean cluster positions served as the template vertices to which all training studies were aligned by rigid and nonrigid coherent point drift (34). As such, a statistical shape model (35) was created, which considers natural deformations present in the data during registration to unseen data. Last, the region label value with highest occurrence at each vertex and the mean positions of each APP landmark were respectively selected as template annotations.

Primitive alignment.— Predefined geometric surrogate primitives were used to facilitate the parameter derivations. This comprised calculating the APP, aligning spheres to the femoral heads, identifying axes through the femoral shafts and necks, as well as fitting a plane through the acetabular margin (Fig 4). Suitable derivation routines relied on the transferred annotations and were implemented in accordance to procedures described in the respective literature (10,16,36).

Geometric parameters.— Hip parameters were derived based on established geometric primitives. The CCD angle was calculated from the femoral axes. For the CE angle, the superior maximal roofage was identified in the proximity of an intersection point identified by the APP placed through the femoral head center. The acetabular depth was measured from the head center to the acetabular margin plane, as done in Pfirrmann et al (37). The inclination was measured from the acetabular

06:00 to the 12:00 position. The anteversion was measured on the axial plane, established by the APP, going through the femoral head center. For the alpha angle, the femoral 12:00, 01:30, and 03:00 o'clock positions were reported. The femoral 12:00 position was defined similarly to that in prior work (38). For each clock position, a bounding circle was fitted to the head sphere to identify the alpha angle in the corresponding plane. The HNO and its ratio were measured at the femoral 03:00 position.

Evaluation.— The performance of the localization network was evaluated by the distance to the ground truth position. The segmentation performance was measured by the Dice similarity coefficient for overlap, the average symmetric surface distance (ASSD) reflecting delineation performance, and the Hausdorff distance (HD) indicating the presence of larger deviations. Annotation transfer was evaluated for respective individuals by region label accuracy (ie, correct vertex category prediction and landmark deviations from manual placed APP points). Primitive alignment was assessed by angular difference and distance to manual annotations. The hip parameters were assessed by comparison with the manual measurements.

Statistical Analysis

The evaluation metrics were calculated per fold and averaged across each fold. To give a detailed picture, values were reported as average mean $\pm$ average standard deviation across all folds followed by the lowest (L) and highest (H) fold mean values in brackets. A Bonferroni corrected significance level of .005 (.05/10) was used for the *P* values of the two-sided paired-sample *t* tests. In addition, mean difference and 10th and 90th percentiles, as well as the intraclass correlation coefficient (ICC)

Figure 4: Fitting procedures used for subsequent parameter derivation: **(A)** anterior pelvic plane (APP) and its contact points and normal n_{APP} **(B)** acetabular margin plane with contact points of parameter measurements and the acetabular 6:00 position, and **(C)** femoral shaft axis and its primitive cylinder, the neck axis, as well as a bounding sphere to the femoral head.

Table 1: Quantitative Evaluation of Automated Bone Segmentation

Label	DSC (%)		ASSD (mm)		HD (mm)	
	Mean	Fold Range	Mean	Fold Range	Mean	Fold Range
Right femur	98.13 ± 0.41	98.05–98.20	0.33 ± 0.06	0.31–0.34	3.10 ± 1.26	2.90–3.37
Left femur	97.94 ± 0.46	97.72–98.05	0.35 ± 0.06	0.34–0.40	3.39 ± 1.22	3.14–3.89
Right pelvis	96.97 ± 0.47	96.75–97.31	0.34 ± 0.04	0.31–0.37	7.57 ± 2.41	6.90–8.52
Left pelvis	97.03 ± 0.74	96.80–97.27	0.33 ± 0.06	0.32–0.36	5.91 ± 2.29	5.29–6.25
Overall mean	97.52 ± 0.46	97.33–97.70	0.34 ± 0.05	0.32–0.37	4.99 ± 1.03	4.64–5.25

Note.—Values are shown as mean ± standard deviation and fold ranges. ASSD = average symmetric surface distance, DSC = Dice similarity coefficient, HD = Hausdorff distance.

for two-way mixed single measures seeking consistency, were provided. For a robust measure of difference dispersion, the median absolute deviation (MAD) was included. Quantitative agreement of single hip parameter measurements was reported by Bland-Altman plots.

The statistical analyses were performed with the help of third-party packages (NumPy v1.19.0, SciPy v1.5.0, pingouin v0.3.6). The visualization of the hip parameters was performed by using the VTK 9.0.0 package. The Bland-Altman plots were generated by a modified version of the pyCompare v1.4.1 package.

Results

Image Processing

Both femur head landmarks were precisely automatically located across all folds with a mean distance of 1.23 mm ± 0.83 (L, 1.10; H, 1.33) (compared with the manual annotation). The highest distance across all participants was 4.12 mm. For the segmentation masks, a mean Dice coefficient of 97.52% ± 0.46 was determined across all femur and pelvis segmentations (Table 1). The lowest Dice coefficient across all participants occurred for the left pelvis with a value of 94.39%.

Similar agreement was conveyed by ASSD values of 0.34 mm ± 0.05 with a highest observed value of 0.56 mm. In addition, a mean HD value of 4.99 mm ± 1.03 indicated only moderate maximum deviations from the ground truth masks. More detailed information for individual bones are shown in Table 1.

Qualitative analysis by S.S.W. revealed no major mask errors were made by the model. In accordance, none of the femur masks showed prominent deviations from the ground truth. For the pelvis, three cases showed larger deviations at the anterior inferior iliac spine (HD values of 11.18 mm, 11.87 mm, and 11.31 mm). A fourth prediction contained a hole in the iliac crest (HD of 13.00 mm). In another case, holes occurred in the thin iliac fossa (HD of 10.30 mm). Beyond these cases, only small local segmentation errors at the superior and inferior ramus of the pubis were present. In four cases, the acetabular margin contained minor visible irregularities.

Morphometric Analysis

Shape model.— The template generation and annotation transfer proved robust with a mean accuracy of 95.14% ± 0.65 (L, 93.17%; H, 96.47%) for the three femur regions of

Table 2: Manual Hip Parameter Measurements

Anatomic Parameter	First Reader	Second Reader	Interobserver			
	Mean $\pm$ SD	Mean $\pm$ SD	P Value	Difference	MAD	ICC
CCD angle (degrees)	129.60 $\pm$ 5.52	129.08 $\pm$ 4.70	.25	−0.55 (−5.74, 4.14)	2.67	0.695
CE angle (degrees)	33.10 $\pm$ 8.00	33.97 $\pm$ 7.29	.13	0.87 (−5.19, 7.06)	3.43	0.777
HNO (mm)	9.16 $\pm$ 1.69	9.92 $\pm$ 1.25	<.001	0.76 (−1.21, 2.70)	1.07	0.335
HNO ratio	0.203 $\pm$ 0.034	0.219 $\pm$ 0.022	<.001	0.017 (−0.023, 0.058)	0.024	0.218
Acetabular depth (mm)	6.90 $\pm$ 2.63	8.85 $\pm$ 2.56	<.001	1.96 (−0.83, 5.03)	1.40	0.650
Acetabular anteversion (degrees)	17.62 $\pm$ 4.32	19.04 $\pm$ 4.51	<.001	1.42 (−2.24, 5.36)	2.14	0.755
Acetabular inclination (degrees)	41.18 $\pm$ 4.04	37.70 $\pm$ 3.60	<.001	−3.48 (−7.14, −0.02)	2.00	0.684
Alpha 12:00 (degrees)	47.69 $\pm$ 8.73	48.79 $\pm$ 8.42	.36	−1.10 (−8.60, 9.03)	4.02	0.222
Alpha 01:30 (degrees)	50.49 $\pm$ 8.64	51.84 $\pm$ 4.60	.17	1.35 (−6.22, 9.48)	3.34	0.223
Alpha 03:00 (degrees)	46.27 $\pm$ 7.02	47.99 $\pm$ 4.31	.02	1.73 (−4.69, 9.49)	4.00	0.347

Note.—The difference is shown as mean (10th, 90th percentile). CCD = centrum-collum-diaphyseal, CE = center-edge, ICC = intraclass correlation coefficient, HNO = head-neck offset, MAD = median absolute deviation, SD = standard deviation.

interest (head, neck, shaft) and $87.16\% \pm 2.51$ (L, 79.65%; H, 92.83%) for the acetabular margin comparing the subregions of the eight region-based annotated participants (Fig 2C, 2D) with the automatically generated annotations. These trends also held true for the mesh vertices of the predicted segmentation, which showed a similar mean accuracy of $95.20\% \pm 0.44$ (L, 93.10%; H, 96.70%) and $87.45\% \pm 2.01$ (L, 79.52%; H, 92.95%). The transferred APP points were $4.46 \text{ mm} \pm 2.07$ (L, 4.01; H, 5.36) away from the manual annotation. However, the observed differences were mostly caused by tangential shifts on the surface. As such, the resulting plane normal n_{APP} followed manual annotations closely with a difference angle of $0.93^\circ \pm 0.57$ (L, 0.76° ; H, 1.16°).

Primitive alignment.— The geometric primitives were successfully fit by the employed procedures. For the femora, radii differences of $-0.33 \text{ mm} \pm 0.63$ (L, 0.01 mm; H, 0.57 mm) and distances to the annotated center of $1.37 \text{ mm} \pm 0.56$ (L, 1.16 mm; H, 1.56 mm) were reported. The angle between annotated and derived shaft and neck line was $4.14^\circ \pm 2.23$ (L, 3.27° ; H, 4.97°) and $3.76^\circ \pm 1.81$ (L, 3.17° ; H, 3.71°), respectively. For both axes, the derivation procedure showed a systematic difference between manual annotated and calculated axes. Nonetheless, only minor differences between fitting results on predicted and manual segmentation masks were observed, as was affirmed as not significant ($P > .115$) across all folds.

Geometric parameters.— Parameter measurements of the first and second reader and their interobserver variability are illustrated in Table 2. Most of the collected parameters showed similar values for mean and standard deviation with small differences. Intermediate to high correlations ($\text{ICC} > 0.650$) could be observed for CCD angle, CE angle, acetabular depth, as well as anteversion and inclination. Nonetheless, in six measurements, small but significant differences were detected between the two readers. For the HNO, as well as the HNO

ratio, the ICC was poor, but the MAD deviations remained minor. The ICC values of the alpha angles were low regardless of the clock position (12:00, 0.222; 01:30, 0.223; and 03:00, 0.347). A mean interdecile range of 15.80° indicated disagreements, especially at the 12:00 position (17.63°).

Automatically derived parameters on the manual and predicted masks are included in Table 3. The derived measurements showed overall high agreement ($\text{ICC} > 0.775$) except for the 12:00 alpha angle (ICC , 0.347) and only minor deviations ($\text{MAD} < 2.58^\circ$ and $< 0.57 \text{ mm}$). No significant mean difference was observed for any of the anatomic parameters assessed. The ranges between the 10th and 90th percentiles were narrow. All MAD values were below the interobserver variability.

The agreement between the readers and the parameter derivation on the manual and predicted masks are depicted in Table 4. In both cases, there was a significant deviation of mean values for the CCD angle and CE angle, which were underestimated by the approach. Little difference between the calculations based on manual and predicted masks was observed, except for a slight increment of MAD values. A fair agreement was achieved for the HNO (ICC , 0.417 [manual mask vs manual] and 0.465 [predicted mask vs manual]). The alpha angles showed fair agreement for the 01:30 and 03:00 positions. The 12:00 position measurements varied markedly as with the manual observations, but the MAD ($< 3.09^\circ$ vs 4.02°) and ICC (> 0.393 vs 0.222) were slightly better. The agreement of each single measurement is encompassed in Bland-Altman plots in Figure 5. Few, but strong deviations were observed for the alpha angles, resulting in large 95% limits of agreement.

Discussion

In this study, a fully automated approach for imaging-based hip morphometry was developed and evaluated. Previous studies have focused either on the localization of landmarks, the segmentation of bones, or the derivation of specific geometric parameters with various degrees of automation. We brought these components together and validated an integrated frame-

Table 3: Automated Hip Parameter Measurements

Anatomic Parameter	Manual Mask	Predicted Mask	Variability			
	Mean $\pm$ SD	Mean $\pm$ SD	P Value	Difference	MAD	ICC*
CCD angle (degrees)	125.88 $\pm$ 4.73	126.35 $\pm$ 4.87	.31	0.47 (–1.35, 2.85)	1.00	0.933
CE angle (degrees)	28.45 $\pm$ 7.45	28.89 $\pm$ 7.28	.59	0.44 (–2.53, 4.00)	1.52	0.925
HNO (mm)	9.62 $\pm$ 1.61	9.77 $\pm$ 1.75	.45	0.15 (–0.85, 1.16)	0.57	0.862
HNO ratio	0.207 $\pm$ 0.031	0.208 $\pm$ 0.032	.39	0.003 (–0.018, 0.026)	0.012	0.821
Acetabular depth (mm)	8.37 $\pm$ 2.27	8.43 $\pm$ 2.10	.74	0.05 (–0.74, 0.86)	0.50	0.934
Acetabular anteversion (degrees)	19.43 $\pm$ 5.04	19.25 $\pm$ 4.98	.37	–0.19 (–2.37, 1.71)	1.12	0.920
Acetabular inclination (degrees)	40.96 $\pm$ 4.07	40.65 $\pm$ 3.85	.30	–0.31 (–2.18, 1.26)	0.85	0.847
Alpha 12:00 (degrees)	48.41 $\pm$ 8.71	48.43 $\pm$ 7.74	.42	0.02 (–7.73, 6.17)	2.58	0.347
Alpha 01:30 (degrees)	49.28 $\pm$ 5.40	48.78 $\pm$ 5.58	.41	–0.50 (–3.44, 2.51)	1.70	0.843
Alpha 03:00 (degrees)	46.97 $\pm$ 6.25	46.86 $\pm$ 5.79	.45	–0.11 (–4.08, 4.92)	2.22	0.775

Note.—Difference shown as mean (10th, 90th percentile). CCD = centrum-collum-diaphyseal, CE = center-edge, HNO = head-neck offset, ICC = intraclass correlation coefficient, MAD = median absolute deviation, SD = standard deviation.

* The ICC was calculated on a per-measurement basis (independent of cross-validation folds).

Table 4: Manual versus Automated Hip Parameter Measurements

Anatomic Parameter	On Manual Mask versus Manual				On Predicted Mask versus Manual			
	P Value	Difference	MAD	ICC*	P Value	Difference	MAD	ICC*
CCD angle (degrees)	.002	–3.46 (–7.62, 0.22)	2.20	0.756	.003	–2.99 (–7.93, 0.53)	2.18	0.755
CE angle (degrees)	<.001	–5.09 (–9.00, –1.03)	1.84	0.887	.001	–4.65 (–8.90, –0.67)	2.07	0.858
HNO (mm)	.50	0.08 (–1.75, 2.00)	1.15	0.417	.38	0.23 (–1.43, 2.14)	0.87	0.465
HNO ratio	.50	–0.004 (–0.043, 0.037)	0.026	0.266	.36	–0.001 (–0.039, 0.040)	0.021	0.288
Acetabular depth (mm)	.19	0.50 (–0.65, 1.78)	0.75	0.882	.20	0.55 (–0.77, 1.92)	0.77	0.840
Acetabular anteversion (degrees)	.14	1.10 (–1.40, 4.21)	1.37	0.847	.18	0.92 (–2.15, 4.09)	1.52	0.796
Acetabular inclination (degrees)	.05	1.52 (–1.40, 4.10)	1.56	0.759	.20	1.20 (–1.39, 3.68)	1.39	0.761
Alpha 12:00 (degrees)	.58	0.17 (–5.72, 8.23)	3.09	0.545	.59	0.19 (–7.22, 7.53)	3.05	0.393
Alpha 01:30 (degrees)	.30	–1.88 (–9.16, 4.93)	3.01	0.346	.24	–2.38 (–9.25, 3.26)	3.30	0.347
Alpha 03:00 (degrees)	.21	–0.16 (–6.07, 6.23)	3.51	0.471	.24	–0.27 (–7.02, 5.82)	3.34	0.360

Note.—Difference shown as mean (10th, 90th percentile). CCD = centrum-collum-diaphyseal, CE = center-edge, HNO = head-neck offset, ICC = intraclass correlation coefficient, MAD = median absolute deviation.

* The ICC was calculated on a per-measurement basis (independent of cross-validation folds).

work for a fully automated pipeline for generic MRI hip morphometry on cohort data. The advances of deep learning, recent registration, and further graphical processing algorithms enable the efficient analysis of a growing amount of medical imaging data and avert costly human intervention.

The localization algorithm proposed was on par with the alignment of a bounding sphere on the manual segmentation masks. Excellent mask delineations were achieved by the proposed network architectures that surpass classic machine learning algorithms despite the intricate imaging data and the fat saturation, severe texture inhomogeneities, artifacts, and cystic elements. This result is also reflected in Table 3 where little impact on subsequent shape processing was seen in most cases. The segmentation results from this study are in line with a recent study for femur MRI segmentation with Dice scores between

0.94 $\pm$ 0.05 (22). The annotation transfer functioned reliably. Compared with related work (39), we were able to transfer region as well as landmark annotations. In addition, we did not rely on prevalent characteristics as done by Zhang et al (40).

In comparison with approaches relying on CT, it should be noted that the comparatively low resolution of the acquired MRI dataset (leaving a small margin for error and partial volume effects, resulting in smooth transitions between bone and cartilage) made the bone delineation ambiguous. As such, deviations of the measurements cannot only be attributed to the processing approach itself, but to the imaging data, resulting manual segmentation irregularities, or manual measurements performed on varying oblique axes. Therefore, prominent differences were not only present between the automated and manual approach, but also between the readers.

Figure 5: Bland-Altman plots of manual versus automated geometric hip parameter measurements show differences in agreement against mean values. Solid lines indicate overall mean difference with dotted lines showing the 95% limits of agreement (mean $\pm$ 1.96 standard deviation) and gray areas indicating respective confidence intervals. Oriented triangles mark individuals within different folds. The four blue diamond symbols represent the mean fold performances. α = acetabular, CCD = centrum-collum-diaphyseal, CE = center-edge, HNO = head-neck offset.

We observed deviations especially in areas of poor imaging contrast (eg, the 12:00 alpha position). As such, this study shows that the alpha angle derivation procedure is sensitive to changes in the underlying segmentation mask because the intersection of the bounding circle and the surface contour can vary largely based on the contour shape. Further differences occurred between manual measurements, such as the HNO and its ratio for which the measured variables were on the same order as the imaging resolution.

Besides measurement irregularities and prediction errors, varied aspects contributed to certain interdecile ranges. For measurements in the proximity of the acetabular rim, such as the CE angle, measurement irregularities can occur due to mesh processing and smoothing, leading to erosion of thin edge borders. In addition, the varying height of the cutoff position on the femoral shaft can have an impact on the CCD angle measurements, which is in line with the angular difference seen in comparison with manually drawn femoral axes.

This study had limitations. The sample size was limited due to the annotation cost of all intermediate steps. Throughout the processing steps, we did not account for present pathologic conditions. A detailed comparison with established hip parameter values was omitted and the interested reader is referred to the cited literature. Further, it is beyond the scope to investigate the influence and error propagation of each processing step. As such, measurements can be influenced by varying degree of segmentation masks, shape processing, registration mismatch, and employed derivation procedures.

In all, we showed that by combining learning and non-learning-based algorithms, an accurate automatic analysis of geometric hip parameters could be demonstrated for MRI data of the GNC study. For all subtasks, including deep learning-based segmentation and localization of landmarks, as well as shape model-based annotation transfer, an overall high accuracy and robustness could be shown. Furthermore, we observed reasonable measurement agreement for nine of 10 assessed parameters and overall low median deviations. Advances in deep learning and the increasing availability of larger cohort studies opens the possibility to leverage tissue characteristics and further quantitative and qualitative findings directly from the same participant and sequence. Given the excellent localization of the femur head centers, future approaches suited to cohort data may circumvent complex and cumbersome graphical processing steps and geometric primitive alignment. By relying solely on learning-based procedures, a robust and accurate parameter derivation in the presence of severe adverse influences may be achieved.

Acknowledgments: The project was conducted with data from the German National Cohort (GNC) (www.nako.de). The GNC is funded by the Federal Ministry of Education and Research (BMBF) (project funding reference numbers: 01ER1301A/B/C and 01ER1511D), federal states, and the Helmholtz Association with additional financial support from the participating universities and the institutes of the Leibniz Association. The authors thank all participants who took part in the GNC study and the staff in this research program.

Author contributions: Guarantors of integrity of entire study, M.F., M.N.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors;

literature research, M.F., S.S.W., T.H., M.Z., M.N., F.S.; clinical studies, M.N.; experimental studies, M.F., S.S.W., M.Z.; statistical analysis, M.F., T.H.; and manuscript editing, M.F., S.S.W., T.H., M.N., F.S., B.Y.

Disclosures of Conflicts of Interest: M.F. Activities related to the present article: institution funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 325028047. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. S.S.W. Activities related to the present article: institution funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 325028047. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. T.H. disclosed no relevant relationships. M.Z. Activities related to the present article: institution funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 325028047. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. M.N. Activities related to the present article: institution received grant from German Research Council (grant number NO GZ: NO 1042/1-1 | SCHI 498/10-1 | YA 28/14-1) Automatisierte Segmentierung und Quantifizierung von geometrischen und strukturellen Parametern anhand von 3D-MRTD Datensätzen der Hüftgelenke: Entwicklung zuverlässiger Algorithmen zur Datenanalyse in großen Kohortenstudien (Nationale Kohorte). Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. F.S. disclosed no relevant relationships. B.Y. disclosed no relevant relationships.

References

- Hack K, Di Primio G, Rakhra K, Beaulé PE. Prevalence of cam-type femoroacetabular impingement morphology in asymptomatic volunteers. *J Bone Joint Surg Am* 2010;92(14):2436–2444.
- Zhang C, Li L, Forster BB, et al. Femoroacetabular impingement and osteoarthritis of the hip. *Can Fam Physician* 2015;61(12):1055–1060.
- Sutter R, Dietrich TJ, Zingg PO, Pfirrmann CWA. How useful is the alpha angle for discriminating between symptomatic patients with cam-type femoroacetabular impingement and asymptomatic volunteers? *Radiology* 2012;264(2):514–521.
- Terjesen T, Gunderson RB. Reliability of radiographic parameters in adults with hip dysplasia. *Skeletal Radiol* 2012;41(7):811–816.
- Clohisy JC, Carlisle JC, Trousdale R, et al. Radiographic evaluation of the hip has limited reliability. *Clin Orthop Relat Res* 2009;467(3):666–675.
- Harris MD, Reese SP, Peters CL, Weiss JA, Anderson AE. Three-dimensional quantification of femoral head shape in controls and patients with cam-type femoroacetabular impingement. *Ann Biomed Eng* 2013;41(6):1162–1171.
- Chadayammuri V, Garabekyan T, Jesse MK, et al. Measurement of lateral acetabular coverage: a comparison between CT and plain radiography. *J Hip Preserv Surg* 2015;2(4):392–400.
- Cerveri P, Marchente M, Bartels W, Corten K, Simon JP, Manzotti A. Automated method for computing the morphological and clinical parameters of the proximal femur using heuristic modeling techniques. *Ann Biomed Eng* 2010;38(5):1752–1766.
- Gras F, Gottschling H, Schröder M, Marintschev I, Reimers N, Burgkart R. Sex-specific differences of the infraacetabular corridor: a biomorphometric CT-based analysis on a database of 523 pelves. *Clin Orthop Relat Res* 2015;473(1):361–369.
- Ng KCG, Lamontagne M, Labrosse MR, Beaulé PE. Comparison of anatomical parameters of cam femoroacetabular impingement to evaluate hip joint models segmented from CT data. *Comput Methods Biomech Biomed Imaging Vis* 2016;6(3):293–302.
- Tawada K, Iguchi H, Tanaka N, et al. Is the canal flare index a reliable means of estimation of canal shape? Measurement of proximal femoral geometry by use of 3D models of the femur. *J Orthop Sci* 2015;20(3):498–506.
- Hartel MJ, Petersik A, Schmidt A, et al. Determination of femoral neck angle and torsion angle utilizing a novel three-dimensional modeling and analytical technology based on CT datasets. *PLoS One* 2016;11(3):e0149480.
- Thiesen DM, Prange F, Berger-Groch J, et al. Femoral anteversion-A 3D CT Analysis of 1232 adult femurs. *PLoS One* 2018;13(10):e0204961.
- Lerch TD, Degonda C, Schmaranzer F, et al. Patient-Specific 3-D Magnetic Resonance Imaging-Based Dynamic Simulation of Hip Impingement and Range of Motion Can Replace 3-D Computed Tomography-Based Simulation for Patients With Femoroacetabular Impingement: Implications for Planning Open Hip Preservation Surgery and Hip Arthroscopy. *Am J Sports Med* 2019;47(12):2966–2977.
- Zhang RY, Su XY, Zhao JX, Li JT, Zhang LC, Tang PF. Three-dimensional morphological analysis of the femoral neck torsion angle-an anatomical study. *J Orthop Surg Res* 2020;15(1):192.

16. Hu J, Xu L, Jing M, Zhang H, Wang L, Chen X. An approach to automated measuring morphological parameters of proximal femora on three-dimensional models. *Int J CARS* 2020;15(1):109–118.
17. 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.
18. Gollwitzer H, Suren C, Strüwind C, et al. The natural alpha angle of the femoral head-neck junction: a cross-sectional CT study in 1312 femurs. *Bone Joint J* 2018;100-B(5):570–578.
19. Samim M, Eftekhary N, Vigdorchik JM, et al. 3D-MRI versus 3D-CT in the evaluation of osseous anatomy in femoroacetabular impingement using Dixon 3D FLASH sequence. *Skeletal Radiol* 2019;48(3):429–436.
20. Xia Y, Fripp J, Chandra SS, Walker D, Crozier S, Engstrom C. Automated 3D quantitative assessment and measurement of alpha angles from the femoral head-neck junction using MR imaging. *Phys Med Biol* 2015;60(19):7601–7616.
21. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
22. Deniz CM, Xiang S, Hallyburton RS, et al. Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. *Sci Rep* 2018;8(1):16485.
23. Zeng G, Zheng G. Deep Volumetric Shape Learning for Semantic Segmentation of the Hip Joint from 3D MR Images. In: *International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging*. Cham, Switzerland: Springer, 2018; 35–48.
24. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med* 2018;79(4):2379–2391.
25. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12(3):e1001779.
26. Bamberg F, Kauczor HU, Weckbach S, et al. Whole-body MR imaging in the German national cohort: Rationale, design, and technical background. *Radiology* 2015;277(1):206–220.
27. Hepp T, Fischer M, Winkelmann MT, et al. Fully Automated Segmentation and Shape Analysis of the Thoracic Aorta in Non-contrast-enhanced Magnetic Resonance Images of the German National Cohort Study. *J Thorac Imaging* 2020;35(6):389–398.
28. Mehta S, Rastegari M, Caspi A, Shapiro L, Hajishirzi H. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018; 561–580.
29. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018; 3–19.
30. Papandreou G, Zhu T, Kanazawa N, et al. Towards accurate multi-person pose estimation in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017; 3711–3719.
31. Chen R, Ma Y, Chen N, Lee D, Wang W. Cephalometric Landmark Detection by Attentive Feature Pyramid Fusion and Regression-Voting. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019; 873–881.
32. De Brabandere B, Jia X, Tuytelaars T, Van Gool L. Dynamic filter networks. In: *Advances in Neural Information Processing Systems*, 2016; 667–675.
33. Lawin FJ, Danelljan M, Khan FS, Forsen PE, Felsberg M. Density Adaptive Point Set Registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018; 3829–3837.
34. Myronenko A, Song X. Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell* 2010;32(12):2262–2275.
35. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. *Med Image Anal* 2009;13(4):543–563.
36. Higgins SW, Spratley EM, Boe RA, Hayes CW, Jiranek WA, Wayne JS. A novel approach for determining three-dimensional acetabular orientation: results from two hundred subjects. *J Bone Joint Surg Am* 2014;96(21):1776–1784.
37. Pfirrmann CWA, Mengiardi B, Dora C, Kalberer F, Zanetti M, Hodler J. Cam and pincer femoroacetabular impingement: characteristic MR arthrographic findings in 50 patients. *Radiology* 2006;240(3):778–785.
38. Mascarenhas VV, Rego P, Dantas P, Gaspar A, Soldado F, Consciência JG. Cam deformity and the omega angle, a novel quantitative measurement of femoral head-neck morphology: a 3D CT gender analysis in asymptomatic subjects. *Eur Radiol* 2017;27(5):2011–2023.
39. Schröder M, Gottschling H, Reimers N, Hauschild M, Burgkart R. Automated Morphometric Analysis of the Femur on Large Anatomical Databases with Highly Accurate Correspondence Detection. *Open Med J* 2014;1(1):15–22.
40. Zhang J, Malcolm D, Hislop-Jambrich J, Thomas CDL, Nielsen PMF. An anatomical region-based statistical shape model of the human femur. *Comput Methods Biomech Biomed Eng Imaging Vis* 2014;2(3):176–185.