European Journal of Radiology 177 (2024) 111588

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Automatic quantification of scapular and glenoid morphology from CT scans using deep learning

Osman Berk Satir [a], Pezhman Eghbali [b], Fabio Becce [c], Patrick Goetti [d], Arnaud Meylan [d], Kilian Rothenbühler [c], Robin Diot [d], Alexandre Terrier [b,d], Philippe Büchler [a,*]

[a] ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland
[b] Laboratory of Biomechanical Orthopedics, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[c] Department of Diagnostic and Interventional Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[d] Department of Orthopedics and Traumatology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

ARTICLE INFO

Keywords:
Computed tomography
Deep learning
Morphometry
Osteoarthritis
Shoulder

ABSTRACT

Objectives: To develop and validate an open-source deep learning model for automatically quantifying scapular and glenoid morphology using CT images of normal subjects and patients with glenohumeral osteoarthritis.
Materials and Methods: First, we used deep learning to segment the scapula from CT images and then to identify the location of 13 landmarks on the scapula, 9 of them to establish a coordinate system unaffected by osteoarthritis-related changes, and the remaining 4 landmarks on the glenoid cavity to determine the glenoid size and orientation in this scapular coordinate system. The glenoid version, glenoid inclination, critical shoulder angle, glenopolar angle, glenoid height, and glenoid width were subsequently measured in this coordinate system. A 5-fold cross-validation was performed to evaluate the performance of this approach on 60 normal/non-osteoarthritic and 56 pathological/osteoarthritic scapulae.
Results: The Dice similarity coefficient between manual and automatic scapular segmentations exceeded 0.97 in both normal and pathological cases. The average error in automatic scapular and glenoid landmark positioning ranged between 1 and 2.5 mm and was comparable between the automatic method and human raters. The automatic method provided acceptable estimates of glenoid version ($R^2 = 0.95$), glenoid inclination ($R^2 = 0.93$), critical shoulder angle ($R^2 = 0.95$), glenopolar angle ($R^2 = 0.90$), glenoid height ($R^2 = 0.88$) and width ($R^2 = 0.94$). However, a significant difference was found for glenoid inclination between manual and automatic measurements ($p < 0.001$).
Conclusions: This open-source deep learning model enables the automatic quantification of scapular and glenoid morphology from CT scans of patients with glenohumeral osteoarthritis, with sufficient accuracy for clinical use.

1. Introduction

Quantifying bone morphology by image analysis is an important but time-consuming part of daily clinical practice in musculoskeletal imaging, as it provides valuable information for the diagnosis and management of numerous bone and joint disorders, including in the shoulder. Glenoid morphometric parameters such as glenoid version and inclination have been shown to play a critical role in the surgical planning of total shoulder arthroplasty (TSA), especially in the treatment of glenohumeral osteoarthritis, in which the disease significantly alters bone morphology [1–3]. Other morphological parameters of the glenohumeral joint, such as the critical shoulder angle, have been associated with rotator cuff pathologies [4,5], whereas the glenopolar angle has been used as an outcome predictor in the management of scapular fractures [6].

Shoulder imaging plays a pivotal clinical role not only in the diagnosis and surgical planning of glenohumeral osteoarthritis, but also in evaluating the outcome of TSA [7,8]. Therefore, scapular and glenoid bone morphometry has become of interest. Glenoid version and inclination were originally measured on 2D CT images [9,10]. However, the reliability of 2D methods is questionable as they may depend on the orientation of the CT slice [11] or ignore the 3D characteristics of glenoid geometry [12]. Hence, 3D measurements have been developed. Moineau et al. proposed semi-automatic segmentation of the scapula

* Corresponding author at: ARTORG Center for Biomedical Engineering Research, University of Bern, Freiburgstrasse 3, CH-3010 Bern, Switzerland.
E-mail address: philippe.buechler@unibe.ch (P. Büchler).

https://doi.org/10.1016/j.ejrad.2024.111588
Received 19 January 2024; Received in revised form 28 May 2024; Accepted 24 June 2024
Available online 25 June 2024

Table 1
Patient characteristics in the CT dataset by shoulder pathology.

	Overall (n = 116)	Normal (n = 60)	Pathological (n = 56)
Sex	60 male, 56 female	40 male, 20 female	20 male, 36 female
Average age (years)	65.3	59.7	71.4
Age range (years)	18–88	18–84	36–88
Average body mass index (kg/m^2)	27.1	25.5	28.9
Body mass index range (kg/m^2)	16.9–47.1	17.9–47.1	16.9–41.6
Left/right shoulders	37/79	14/46	23/33
Diagnoses			• 34 primary glenohumeral osteoarthritis (4 with secondary cuff tear) • 17 cuff tear arthropathy • 3 secondary glenohumeral osteoarthritis • 1 fracture (acute and sequelae) • 1 septic arthritis

and calculated glenoid version based on manually identified glenoid surface [13]. Terrier et al. subsequently proposed using curvature information to manually define specific landmarks, and the glenoid surface to determine a coordinate system for measuring glenoid version and inclination [14]. The main limitations of these approaches are the time-consuming and repetitive nature of manual segmentation and landmark positioning, and the dependence on the experience of the human assessor. Commercially available software (e.g., Blueprint (Stryker, Bloomington, MN, USA)) have been developed to assist orthopedic surgeons in surgical planning by automatically measuring clinically relevant anatomical parameters such as glenoid version and inclination [15]. However, no precise information on the automatic measurement algorithm is provided.

This study aimed to investigate the feasibility and reliability of deep learning for automatically quantifying clinically relevant anatomical parameters in the shoulder. Here, we present and validate each step in developing a fully automated pipeline to quantify scapular and glenoid morphology from clinical CT scans. Our approach involves scapula segmentation and landmark localization to define a coordinate system and measure several morphological parameters of the glenoid cavity and acromion.

2. Materials and methods

2.1. Dataset

We retrospectively reviewed 278 patients with glenohumeral osteoarthritis, including cuff tear arthropathy, and 140 trauma patients with normal scapulae between February 2008 and September 2015. Only patients older than 18 years with a CT scan that completely covered (in the transverse and longitudinal planes) one of the two scapulae were included. Patients in whom at least one scapula was not completely included in the reconstructed field-of-view or without a CT image series reconstructed using a smooth kernel were excluded. Our final study dataset consisted of 56 preoperative pathological/osteoarthritic shoulder CT scans from 56 patients with glenohumeral osteoarthritis including cuff tear arthropathy and 60 normal/non-osteoarthritic scapulae from 60 trauma patients. The patient characteristics and types of shoulder pathology included in the study are listed in Table 1. This retrospective study was approved by the institutional ethics committee (CER-VD protocol 2020-01895), and all CT scans were de-identified before image analysis.

CT scans were performed with several generations of multidetector CT systems (8- to 256-detector rows, Lightspeed Ultra, Lightspeed VCT, Discovery CT750 HD, and Revolution CT, all from GE Healthcare). The settings for data acquisition were: tube potential 120 kVp, tube current between 100 mA and 450 mA, gantry revolution time between 0.6 s and 1 s, pitch between 0.875 and 1.375, beam collimation from 8 × 1.25 mm to 256 × 0.625 mm. The relevant image reconstruction parameters were: field-of-view between 18 × 18 cm and 29 × 29 cm, slice thickness between 0.625 mm and 1.25 mm, and smooth reconstruction kernel.

2.2. Scapula segmentation

A musculoskeletal radiologist (12 years of experience) manually segmented all scapulae for training and validation of the deep learning model. We used the nnUNet [16] framework, which follows a U-Net architecture for segmentation and uses 3D patches extracted from the CT scans for training. Parameters relevant to training the segmentation network are listed in Table 2. The accuracy was evaluated with a 5-fold cross-validation comparing the Dice similarity coefficient and Hausdorff distance between automatic and manual segmentations. To estimate the inter-rater variability, two additional human raters manually segmented the scapula on a small subset of our dataset (5 normal and 5 pathological CT scans), and Dice overlap coefficients between the different raters were measured.

Table 2
Summary of the main characteristics of the automatic segmentation and landmarking networks.

	Segmentation	Landmarking
Training	3D volume patches	Full 3D volume
Network	U-Net [16]	Modified U-Net [18]
Optimizer	Stochastic gradient descent	Stochastic gradient descent
Learning rate	0.01	0.001
Dropout	None	None
Number of epochs	400	50
Patch size (pixel)	160 x 160 x 96	128 x 128 x 128
Voxel size (mm)	0.39 x 0.39 x 0.7	1.36 x 1.36 x 1.8
Data augmentation	Random 3D rotation ([-30°,30°]) Random scaling ([0.8–1.2])	Random 3D rotation ([-40°,40°] in mediolateral axis, [-10°, 10°] in other axes)
Loss function	Dice loss + Cross-entropy loss (Equally weighted)	Wing loss [19]

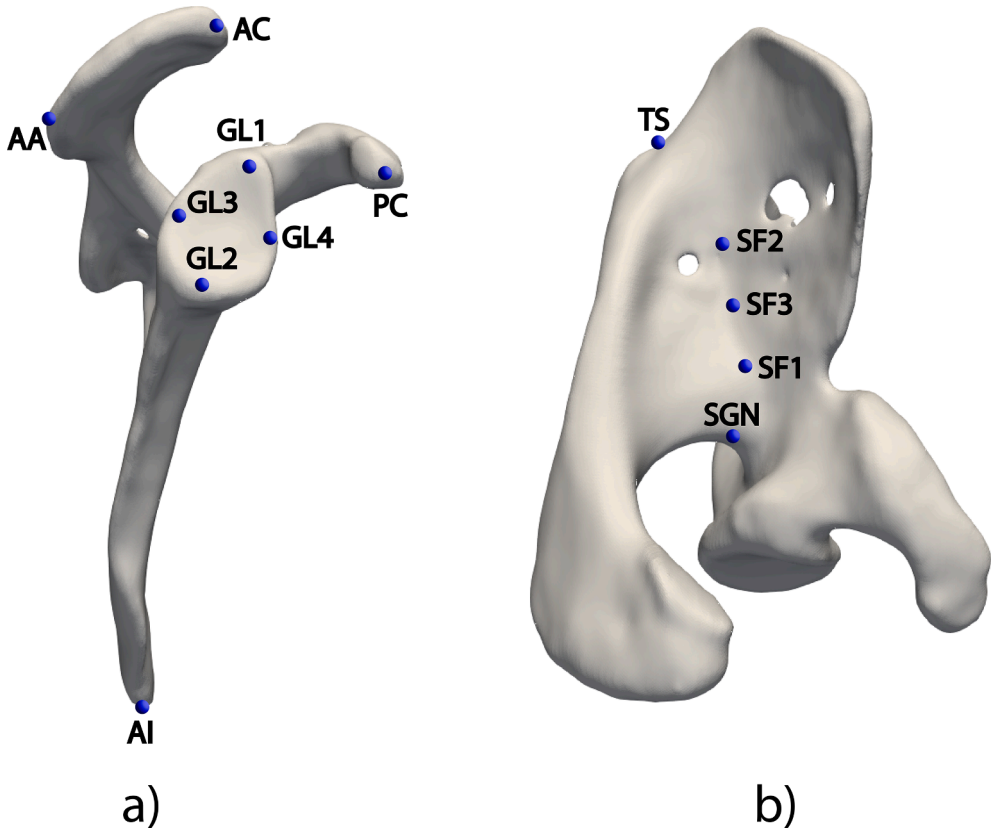

Fig. 1. Location of the 13 landmarks in the lateral (a) and superior (b) view of the scapula. These landmarks were defined at the angulus inferior (AI), the trigonum spinae (TS), the coracoid process (PC), the acromion (AC), the angulus acromialis (AA), the spinoglenoid notch (SGN), three landmarks forming a line at the bottom of the supraspinatus fossa (SF1-SF3), and four landmarks at the edges of the glenoid cavity (GL1-GL4).

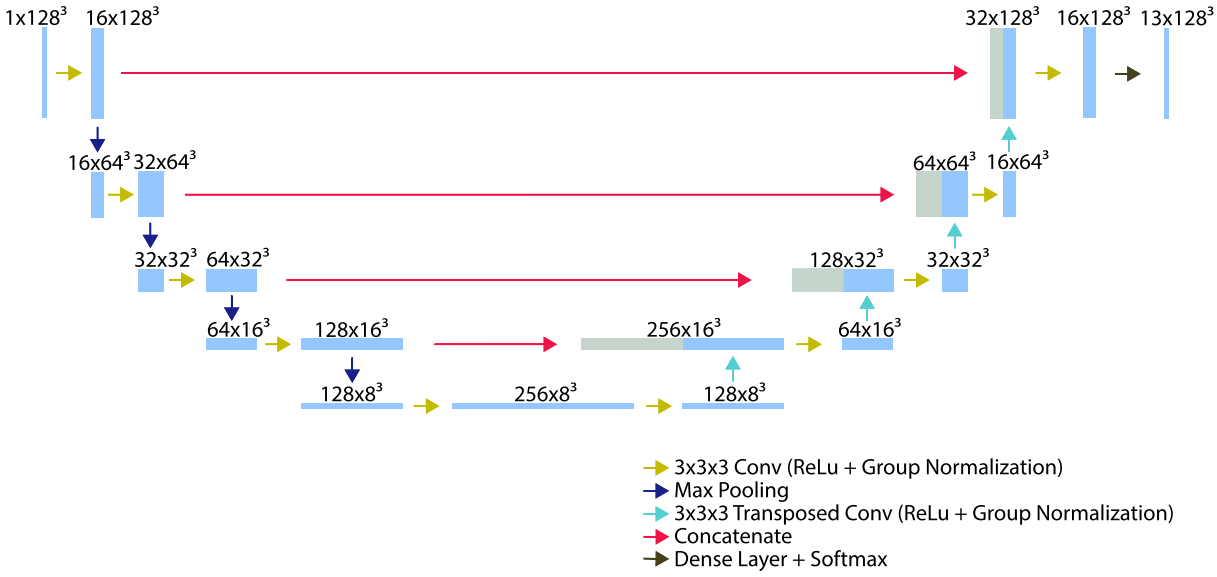

Fig. 2. The modified U-Net [18] architecture used for scapula and glenoid landmark localization.

2.3. Localization of scapular and glenoid landmarks

Nine scapular landmarks were manually placed in 3D Slicer [17] on multiplanar CT reformats, with access to segmented scapular surface meshes to assist human raters in positioning bone landmarks more quickly during manual annotations. These landmarks were used to establish the reference coordinate system of the scapula. Four additional landmarks were placed at the edges of the glenoid cavity to define its height and width (GL1-GL4). Each of these 13 landmarks (Fig. 1) was positioned by 3 human experts (two radiologists and one orthopedic

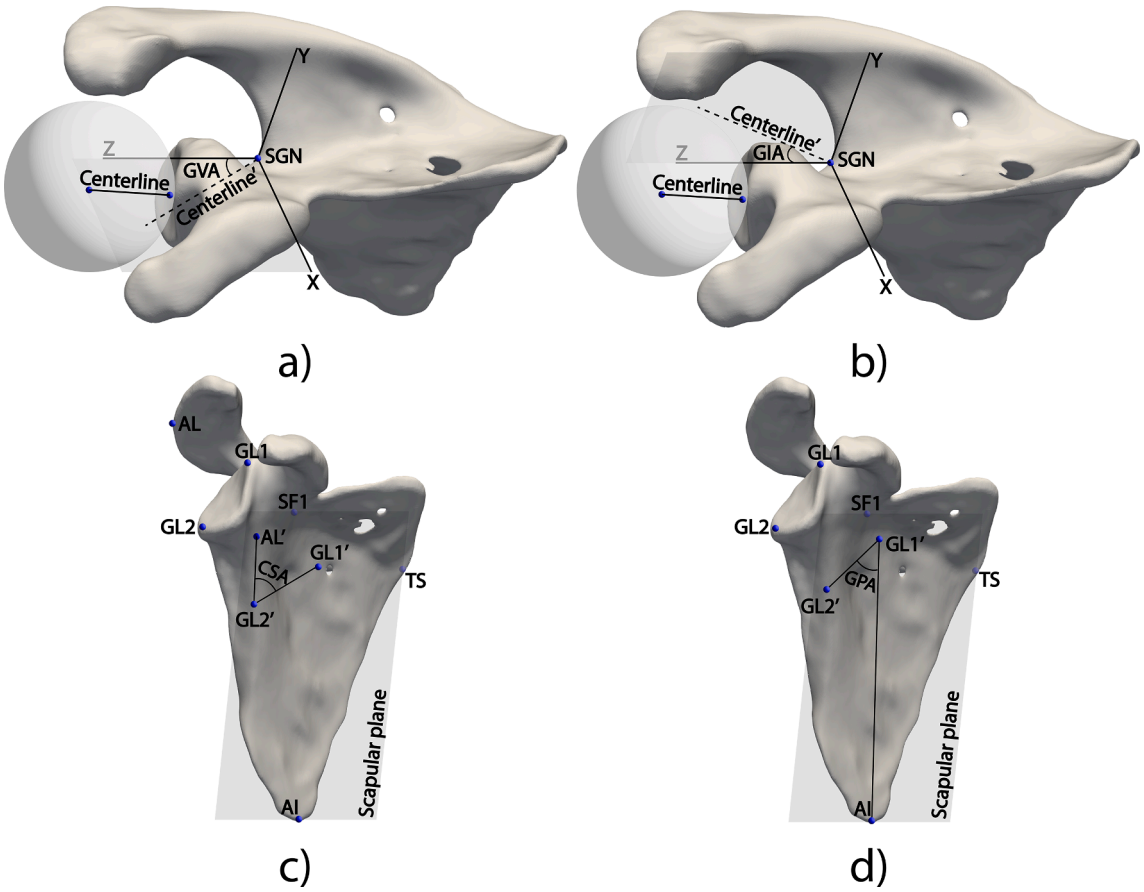

Fig. 3. A) the scapular coordinate system with the glenoid version (GVA), b) the glenoid inclination (GIA), c) the critical shoulder angle (CSA), and d) the glenopolar angle (GPA). the points denoted by an apostrophe indicate the projection of the given point onto the corresponding plane.

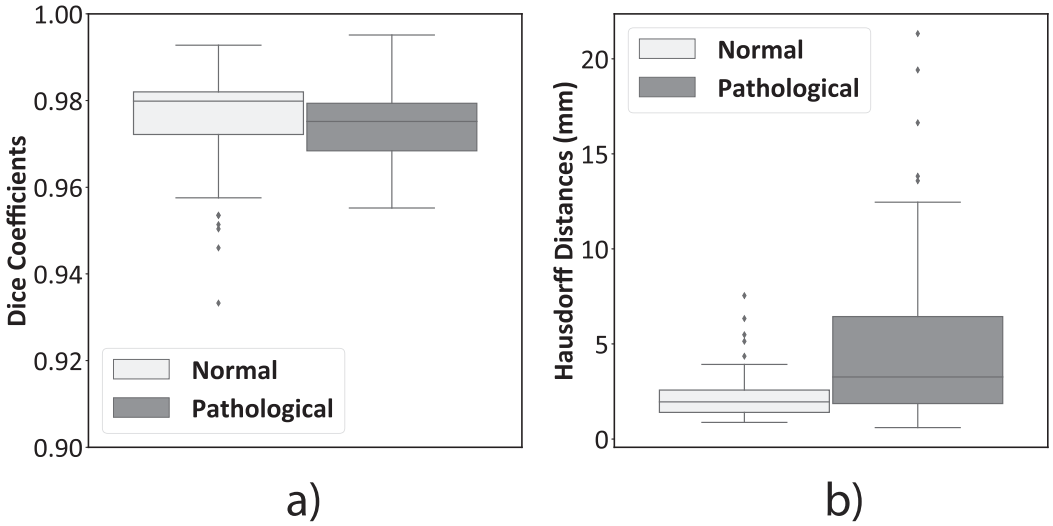

Fig. 4. Dice similarity coefficients (a) and Hausdorff distances (b) between the automatic scapula segmentation and ground truth segmented by the expert human rater.

surgeon).

The reference position of each landmark was determined using a weighted average of these landmarks, with the annotations that differed more from the others receiving less weight. The landmark reference positions were used to train a slightly modified 3D U-Net [18] that predicted landmark positions from binary images of the segmented scapula (Fig. 2). For each landmark, the network generated a 3D scalar field where the value associated with each voxel corresponded to the probability that it is the position of the landmark. The position of each landmark was then determined by the weighted sum of the voxel

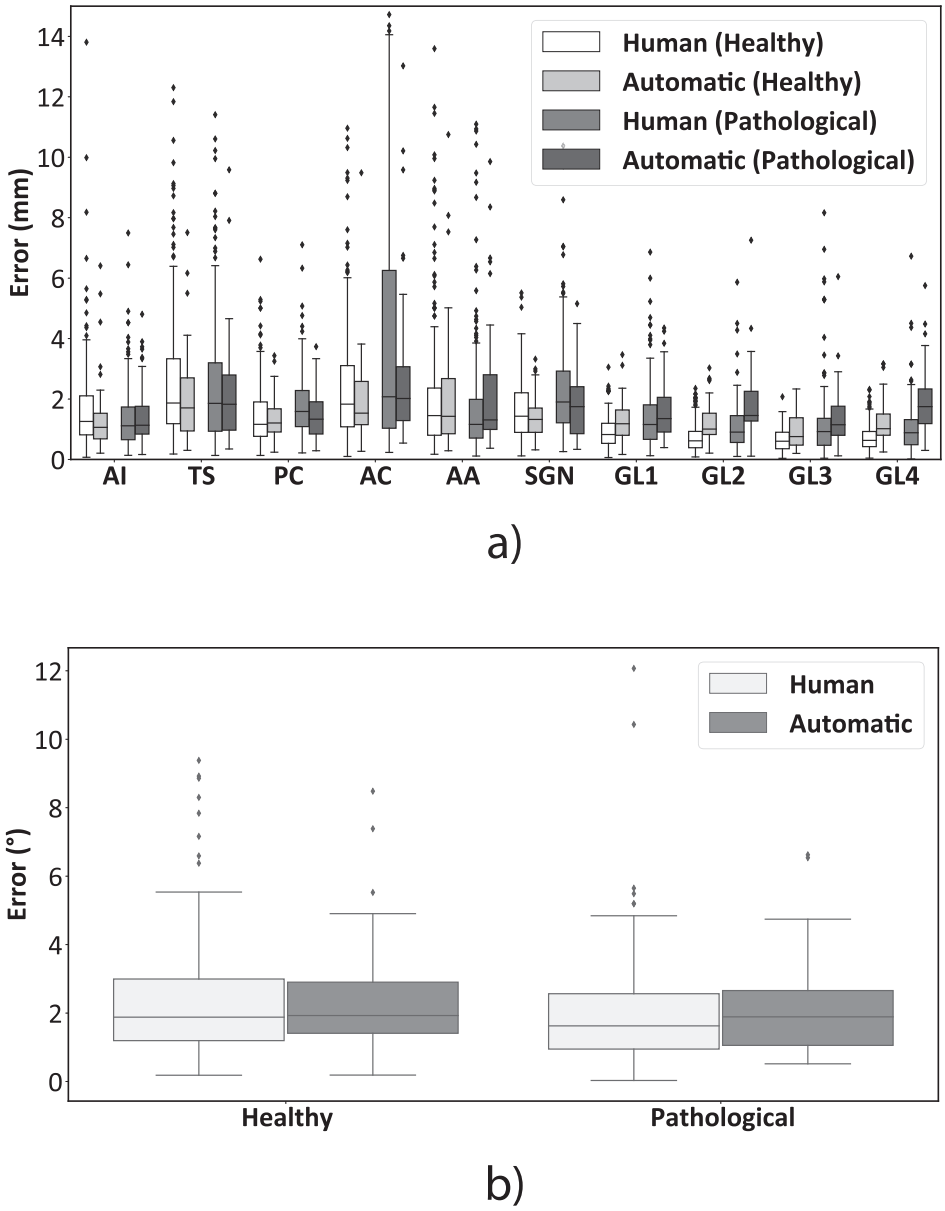

Fig. 5. Comparison between automatic quantification error and inter-rater error for the defined scapular and glenoid landmarks (a), and for the lines formed by landmarks at the supraspinatus fossa (b). The angle error is defined by the angle between the line formed by reference landmarks in the supraspinatus fossa to each human rater and automatic quantification.

Table 3
Summary of the p-values from the comparison between the automatic quantification error and inter-rater error for scapular and glenoid landmark positioning, along with the mean absolute error ± standard deviations from reference landmarks.

	p-values		Difference (mm)			
	Normal	Pathological	Normal		Pathological	
			Prediction	Inter-rater	Prediction	Inter-rater
AI	0.120	0.507	1.34 ± 1.14	1.66 ± 1.50	1.48 ± 1.03	1.38 ± 1.03
TS	0.042	0.239	2.04 ± 1.43	2.66 ± 2.22	2.16 ± 1.69	2.53 ± 2.19
PC	0.307	0.036	1.32 ± 0.65	1.46 ± 1.04	1.49 ± 0.82	1.81 ± 1.07
AC	0.087	0.078	1.96 ± 1.31	2.43 ± 2.00	3.08 ± 3.64	4.44 ± 5.41
AA	0.602	0.465	2.31 ± 2.61	2.59 ± 3.80	2.47 ± 2.83	2.11 ± 3.40
SGN	0.046	0.051	1.39 ± 0.66	1.66 ± 0.99	1.86 ± 1.12	2.30 ± 1.58
GL1	<0.001	0.203	1.22 ± 0.63	0.90 ± 0.52	1.64 ± 0.95	1.43 ± 1.14
GL2	<0.001	<0.001	1.20 ± 0.63	0.72 ± 0.47	1.78 ± 1.12	1.08 ± 0.81
GL3	<0.001	0.282	0.91 ± 0.57	0.66 ± 0.38	1.39 ± 0.94	1.19 ± 1.23
GL4	<0.001	<0.001	1.23 ± 0.67	0.74 ± 0.45	1.91 ± 1.06	1.03 ± 0.84
			Difference (°)			
SF	0.984	0.486	2.3 ± 1.6	2.3 ± 1.6	2.1 ± 1.3	2.0 ± 1.5

coordinates with probabilities from the scalar field. The training loss function was the sum of the Wing losses [19] calculated from the distance between the predicted and reference positions of all landmarks.

The evaluation of the automatic landmark detection included a 5-fold cross-validation in which the distance between the predicted landmarks and the reference landmarks was evaluated. However, since multiple landmarks were used to determine the orientation of the supraspinatus fossa, the detection quality of these landmarks was evaluated based on the angle of the best-fit line and not on the point-wise error. 5 % of the total dataset was used to determine the best-performing network parameters and a 10-fold data augmentation was performed by applying a random 3D rotation (Table 2).

2.4. Calculation of clinically relevant scapular and glenoid anatomical parameters

A scapular reference coordinate system was created for each patient using landmarks [14]. The scapular plane was defined as the plane formed by AI, TS, and SF1, the most lateral landmark in the supraspinatus fossa. Then, SF2 and SF3 were projected onto this plane, and the line that best fits these projections and SF1 defined the scapular axis (Z-axis). The X-axis is orthogonal to the scapular plane, and the Y-axis is orthogonal to the X- and Z-axes (Fig. 3).

A triangular surface mesh was generated from the segmentation using the flying edges algorithm [20], from which the glenoid surface was automatically identified from this mesh using glenoid landmarks. A sphere was fitted to the identified glenoid cavity, and the glenoid centerline was defined as the line connecting the center of gravity of the glenoid surface to the center of the sphere. Clinically relevant anatomical parameters of the scapula and glenoid such as the glenoid version, glenoid inclination, critical shoulder angle and glenopolar angle were automatically calculated, as well as the glenoid height and width (Fig. 3). In this study, the glenoid version and inclination followed the definition of Terrier et al. [14]. The critical shoulder angle was defined as the angle between the infero-superior line of the glenoid (GL1-GL2) and the line from GL2 to the outermost lateral point of the acromion (AL) after these points were projected onto the scapular plane (Fig. 3). AL was determined by projecting the points of the scapula located within 3 mm of the acromion (AC) onto the scapular plane. Of these projections, the most lateral point was selected as AL. The glenopolar angle was defined as the angle between the infero-superior line of the glenoid and the line from AI to GL1 after GL1 and GL2 were projected onto the scapular plane (Fig. 3). In addition to these parameters, we calculated the height and width of the glenoid as the Euclidean distance between GL1-GL2 and GL3-GL4, respectively. The glenoid surfaces were also manually extracted by the same experienced musculoskeletal radiologist for manual measurements.

The automatic measurements were compared with the manual measurements using a five-fold cross-validation of the entire dataset. Statistical comparisons between automatic and manual quantifications were performed using a two-tailed *t*-test, with the significance level set at p-value < 0.05. For the anatomical parameters, the 95 % confidence interval (CI) of the difference between the two methods was also calculated using paired-samples t-tests.

The source code for the methods described in this work is open-source, making it accessible for further development and use by the broader medical and research community. The code is freely available and can be downloaded from the following repository: https://gitlab.epfl.ch/publication/automatic-scapular-morphology.

3. Results

3.1. Scapula segmentation

The 5-fold cross-validation over the entire dataset revealed an excellent average Dice similarity coefficient of 0.97 ± 0.01 between the automatic and manual scapula segmentations for both normal and pathological cases (Fig. 4). The average Hausdorff distances were 2.3 mm and 5.1 mm for normal and pathological shoulders, respectively, demonstrating that the maximum segmentation error was higher for osteoarthritic cases. The regions of higher error were the glenoid rim, the outer aspects of the acromion and coracoid, and the supraspinatus fossa. The Dice coefficients between the different raters varied between 0.94 and 0.99, indicating high inter-rater agreement.

3.2. Localization of scapular and glenoid landmarks

The automatic identification revealed an average deviation between 1.0 mm and 2.5 mm from the reference for all landmarks (Fig. 5). In normal subjects, no statistical differences were found between human raters and the automatic identification of scapular landmarks, except for TS ($p = 0.042$) and SGN ($p = 0.047$). For pathological cases, the difference between human raters and the automatic identification was significant ($p = 0.036$) only for PC. For these three landmarks, the inter-rater error was greater than the automatic error. However, the inter-rater error was significantly smaller than the automatic error for the glenoid landmarks for both normal and pathological cases ($p < 0.001$). Except for the glenoid landmarks, our automatic method demonstrated

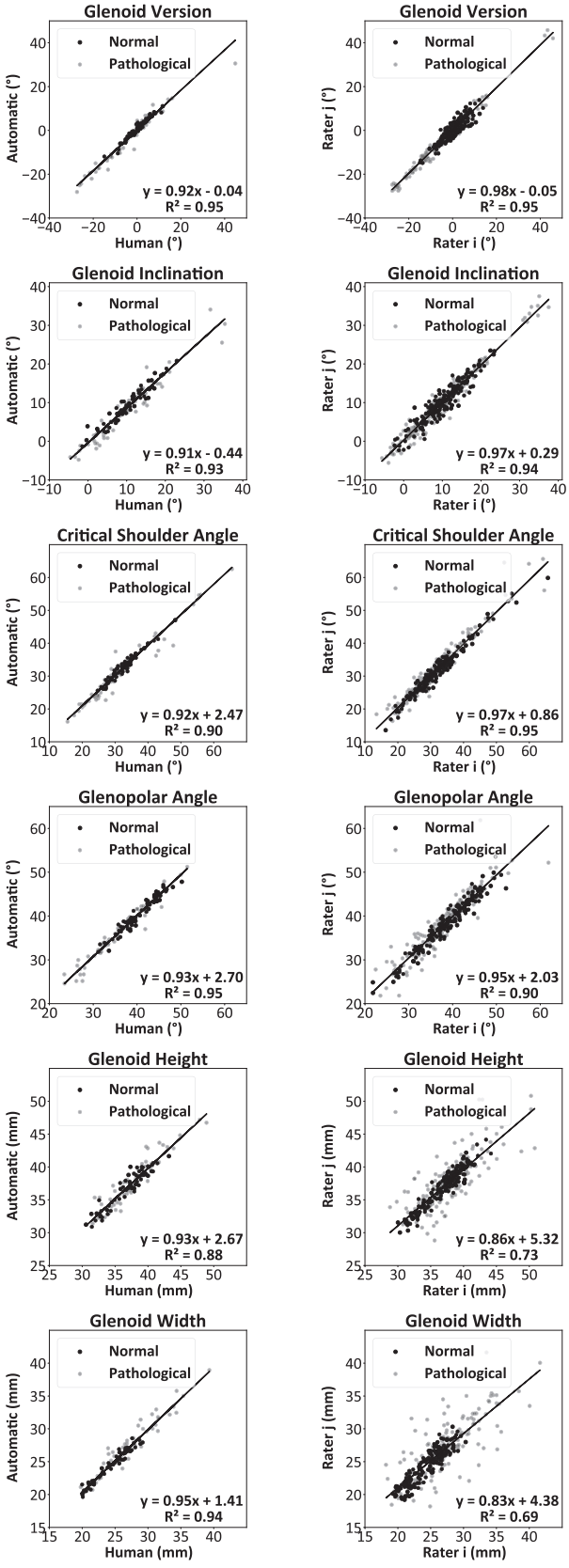

Fig. 6. Comparison between manual and automatic measurements (left) of the anatomical parameters of the scapula and glenoid (glenoid orientation and size), along with the reliability of measurements between different human raters (right).

similar or less error compared to the error between human raters (Table 3).

The average inter-rater error of the supraspinatus fossa line angle was 2.13°±1.58°, while the average automatic quantification error was 2.20°±1.46° (Table 3). No statistical differences were found between the inter-rater and automatic identification error for both normal and pathological cases ($p = 0.984$ and $p = 0.486$, respectively).

3.3. Calculation of clinically relevant scapular and glenoid anatomical parameters

Regression between manually and automatically determined parameters showed good to excellent agreement (Fig. 6). The best-fit slopes were between 0.91 and 0.95, and intercepts were between −0.44° and 2.70°. The coefficients of determination (R^2) between the manually and automatically quantified parameters varied from 0.88 to 0.95 (Table 4). The 95 % CI of the mean difference between manually and automatically quantified anatomical parameters was found to be within less than 1° and 0.5 mm, except for glenoid inclination (Table 5). Figs. 6 and 7 summarize the agreement between automatic and manually quantified anatomical parameters, along with the inter-rater agreement with correlation and Bland-Altman plots, respectively.

No significant differences were found between the manually and automatically quantified anatomical parameters (Table 4), except for glenoid inclination ($p < 0.001$). The high overlap between the 95 % CI of the difference between reference and automatic methods, and between different raters also demonstrated that the difference between automatic and manual quantification is comparable to the differences between human raters, except for glenoid inclination (Table 5). When considering the difference between the manual and automatic glenoid inclination, it was found that the average difference was 1.6°, indicating that the automatic quantification tended to result in a slightly lower glenoid inclination. Nevertheless, the correlation between the manual and automatic glenoid inclination was high, with a regression slope of 0.91 and an R^2 value of 0.93.

Inter-rater agreement was lowest for glenoid height and width, with R^2 values of 0.73 and 0.69, respectively. For other anatomical parameters, the inter-rater agreement and the agreement with the manual quantification resulted in similar slopes of the best-fit lines and R^2 values (Table 4).

4. Discussion

The proposed method is an automated method for measuring scapular anatomy using deep learning and image analysis on CT images. The method consists of segmenting the scapula and identifying relevant bone landmarks to calculate various shoulder parameters that are crucial for the diagnosis and planning of shoulder surgery. The method was validated against manual assessment and showed similar accuracy in bone segmentation and landmark positioning, except for the glenoid landmarks. In this case, however, the landmark positioning error is small (about 1.5 mm). More importantly, manual and automatic measurements of glenoid height and width correlated well, with a higher correlation than between different human raters.

Overall, the automatic measurements correlated well with the manual measurements, with very high R^2 values and slopes of the best-fit lines close to 1. Moreover, the agreement between manual and automatic measurements was comparable to the agreement between different human raters, indicating the high reliability of the proposed approach. The proposed automatic method significantly reduces the time required for segmentation and landmarking to measure anatomical parameters, from over an hour for manual measurement to less than 10 minutes. This efficiency could facilitate the broader application of these measurements in the clinical setting. The reliability and objectivity of the automatic method are significant advantages, as they are not influenced by the operator's experience, subjectivity, or fatigue.

Table 4
P-values, R^2 and best-fit line slopes of the agreement between reference and automatic measurements, along with R^2 and best-fit line slopes of the inter-rater agreement.

	p-values	R^2		Slope	
	Reference-DL	Reference-DL	Inter-rater	Reference-DL	Inter-rater
Glenoid version	0.539	0.95	0.95	0.92	0.98
Glenoid inclination	<0.001	0.93	0.94	0.91	0.97
Critical shoulder angle	0.761	0.90	0.95	0.92	0.97
Glenopolar angle	0.068	0.95	0.90	0.93	0.95
Glenoid height	0.964	0.88	0.73	0.93	0.86
Glenoid width	0.374	0.94	0.69	0.95	0.83

Table 5
Average differences between reference and automatic measurements, along with 95% CI of the mean difference between automatic and reference measurements, and between raters.

	Difference (°)		
	Average	95 % CI	
	Reference-DL	Reference-DL	Inter-rater
Glenoid version	−0.12	(−0.51, 0.27)	(−0.22, 0.22)
Glenoid inclination	1.34	(0.99, 1.70)	(−0.18, 0.18)
Critical shoulder angle	−0.07	(−0.53, 0.39)	(−0.19, 0.19)
Glenopolar angle	−0.21	(−0.45, 0.02)	(−0.21, 0.21)
	Difference (mm)		
Glenoid height	0.004	(−0.21, 0.22)	(−0.19, 0.19)
Glenoid width	−0.07	(−0.24, 0.09)	(−0.23, 0.23)

We assessed the agreement between our manual and automatic measurements and compared it with the agreement reported for the Blueprint software [15]. In this work, the best agreement between manual and automatic methods for the glenoid version had a 95 % CI of the limits of agreement (LoA) of [–3.9°,9°] in 60 patients with glenohumeral osteoarthritis. The same analysis on our automatic and manual measurements for 56 pathological cases in our data set yielded a range of [–5.1°,5.6°]. For glenoid inclination, the agreement between automated method and manual quantification yielded a 95 % CI of LoA within [–9°,9.4°], as reported by the authors, while the same analysis with our methods yielded a range of [–2.7°,6°]. This indirect comparison showed that the agreement between our manual and automated measurements was comparable to the agreement reported for Blueprint in the measurement of the glenoid version and inclination. Both evaluations were performed on similar yet distinct datasets of pathological shoulder conditions.

Despite different measurement approaches and datasets, our measurements of glenoid version and inclination agreed well with those reported for Blueprint in pathological shoulders [15]. However, in addition to glenoid version and inclination, our tool also provides automatic measurements of the critical shoulder angle, glenopolar angle, and glenoid width and height. Furthermore, we have released our tool as open-source software so that clinicians can use it with their own datasets and extend it with additional morphometric measurements (https://gitlab.epfl.ch/publication/automatic-scapular-morphology).

In a more recent study that compared two different software that automatically measures glenoid version and inclination (Blueprint and VIP (Arthrex, Naples, FL, USA)) on 63 CT scans of patients with primary glenohumeral osteoarthritis and rotator cuff arthroplasty or failed rotator cuff repair, considerable variability was found [21] (more than 5° in 30 % of cases for glenoid version and almost half of the cases for glenoid inclination), emphasizing the need for an automated and open-source tool to quantify scapular anatomy. In another study, automatic quantification of the critical shoulder angle based on deep learning was proposed for 8467 anteroposterior radiographs [22]. A median difference of about 1° between the automatic and manual critical shoulder angle was found. However, the work was limited to 2D, and no inter-rater analysis was performed. Our automatic quantification of the critical shoulder angle resulted in lower median errors of 0.7° and 0.9° for normal and pathological cases in our dataset, respectively.

External validation of the model was tested by examining the effects of different CT system models on morphometric quantification. 10 of the 116 CT scans in our study were acquired with older generation CT systems (Lightspeed Ultra and Lightspeed VCT, released in 2001 and 2004, respectively) compared to the rest of the dataset (Discovery CT750 HD and Revolution CT, released in 2008 and 2013, respectively). When the models were trained with the newer scans, validation on older scans showed that the results were consistent across old and new scanner models, indicating the robustness of the method. A further analysis was performed to determine whether the sample size was sufficient for training the deep learning models. We created four different training datasets, comprising 23, 46, 69 and 92 cases respectively, and evaluated the model accuracy using 24 cases. It was found that the dataset used was adequate, as increasing the training set from 69 to 92 cases did not significantly change the accuracy of the automatic quantification.

The main limitation of our study concerns the extraction of the glenoid cavity, both manually and automatically due to poorly defined edges. The use of four landmarks to define the glenoid cavity mitigates this problem, as the exact identification of the glenoid is not crucial to obtain reliable results, as it is mainly used to fit to a sphere, and the center of this sphere is used to determine the glenoid orientation. The possible exclusion of the whole scapula from the CT field of view, which could lead to errors in the coordinate system, is another limitation of the study; however, this can be avoided by proper image acquisition. Another limitation of this study is the number and variety of experience and specialties of the human raters who made the manual annotations. Ideally, multiple raters with different experience and specialization would improve the analysis between raters, but the labor-intensive nature of the manual annotations was a limiting factor. We also calculated the Dice overlap coefficients between automatic segmentation and two additional human raters, who each segmented manually 10 scapulae. The Dice coefficients varied between 0.94 and 0.99, demonstrating high agreement between automatic segmentation and different raters, whose annotations were not used in the training or validation of the model. Therefore, it can be said that although the model was trained with only one set of manual segmentations, the generalization of the model is not affected. Finally, the current dataset includes normal subjects and patients with glenohumeral osteoarthritis and cuff tear arthropathy. While the accuracy of the methods still needs to be assessed for other clinical indications, there are no technical reasons preventing wider application.

In conclusion, this study demonstrates the effectiveness of deep learning-based automatic analysis of CT images for accurate morphometric measurements of scapulae. The results agree very well with human measurements and offer significant time savings in increasingly busy clinical workflows, which can be used for clinical tasks where the contribution of human raters is more important (e.g. qualitative assessment or clinical reasoning). The approach can provide radiologists

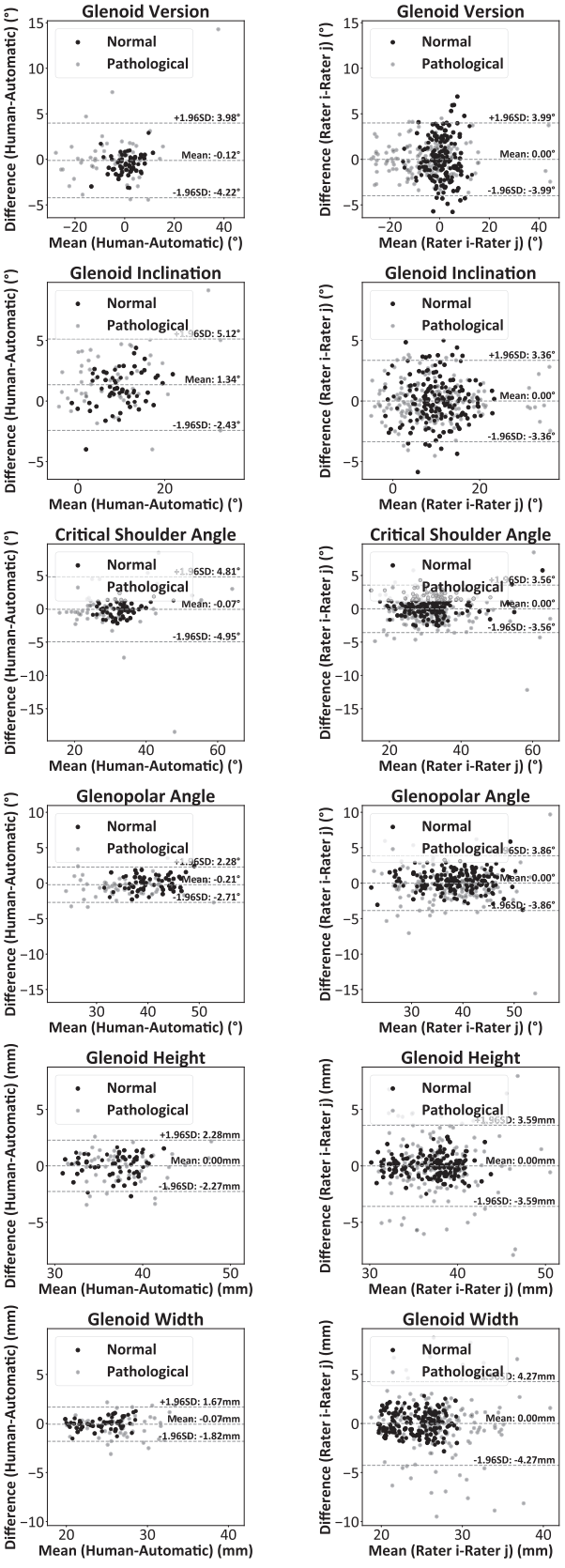

Fig. 7. Bland-Altman plots of the manual and automatic measurements (left) of the anatomical parameters of the scapula and glenoid (glenoid orientation and size), along with the Bland-Altman plots of the different human raters (right).

and surgeons with a reliable understanding of the patient's anatomy, which is crucial for diagnosis and surgical planning. This contributes to informed clinical decision-making and potentially better surgical outcomes. In addition, this method could help to evaluate the relationship between anatomical parameters of the scapula and glenoid and the outcome of a TSA, as it allows for faster and more reliable analysis of larger data sets from clinical databases or registries.

CRediT authorship contribution statement

Osman Berk Satir: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Pezhman Eghbali:** Writing – review & editing, Validation, Software, Methodology. **Fabio Becce:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Patrick Goetti:** Writing – review & editing, Methodology, Investigation, Data curation. **Arnaud Meylan:** Writing – review & editing, Data curation. **Kilian Rothenbühler:** Writing – review & editing, Data curation. **Robin Diot:** Writing – review & editing, Data curation. **Alexandre Terrier:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Philippe Büchler:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Swiss National Science Foundation (grant number: 189972)

References

[1] P. Goetti, P.J. Denard, P. Collin, M. Ibrahim, A. Mazzolari, A. Lädermann, Biomechanics of anatomic and reverse shoulder arthroplasty, EFORT Open Reviews 6 (2021) 918–931, https://doi.org/10.1302/2058-5241.6.210014.

[2] P. Fulin, M. Kysilko, D. Pokorny, R. Padr, N. Kasprikova, I. Landor, A. Sosna, Study of the variability of scapular inclination and the glenoid version - considerations for preoperative planning: clinical-radiological study, BMC Musculoskelet Disord 18 (2017) 16, https://doi.org/10.1186/s12891-016-1381-4.

[3] A. Ganapathi, J.A. McCarron, X. Chen, J.P. Iannotti, Predicting normal glenoid version from the pathologic scapula: a comparison of 4 methods in 2- and 3-dimensional models, J. Shoulder Elbow Surg. 20 (2011) 234–244, https://doi.org/10.1016/j.jse.2010.05.024.

[4] B.K. Moor, S. Bouaicha, D.A. Rothenfluh, A. Sukthankar, C. Gerber, Is there an association between the individual anatomy of the scapula and the development of rotator cuff tears or osteoarthritis of the glenohumeral joint?: A radiological study of the critical shoulder angle, The Bone & Joint Journal 95-B (2013) 935–941, https://doi.org/10.1302/0301-620X.95B7.31028.

[5] Z. Rose-Reneau, A.K. Moorefield, D. Schirmer, E. Ismailov, R. Downing, B. W. Wright, The critical shoulder angle as a diagnostic measure for osteoarthritis and rotator cuff pathology, Cureus (2020), https://doi.org/10.7759/cureus.11447.

[6] J. Romero, P. Schai, A.B. Imhoff, Scapular neck fracture - the influence of permanent malalignment of the glenoid neck on clinical outcome, Arch. Orthop. Trauma Surg. 121 (2001) 313–316, https://doi.org/10.1007/s004020000224.

[7] A. Sharifi, M.J. Siebert, A. Chhabra, How to measure glenoid bone stock and version and why it is important: A practical guide, Radiographics 40 (2020) 1671–1683, https://doi.org/10.1148/rg.2020200008.

[8] J. Weaver, I. Omar, N. Chadwick, J. Shechtel, J. Elifritz, C. Shultz, M. Taljanovic, Update on shoulder arthroplasties with emphasis on imaging, JCM 12 (2023) 2946, https://doi.org/10.3390/jcm12082946.

[9] M.D. Budge, G.S. Lewis, E. Schaefer, S. Coquia, D.J. Flemming, A.D. Armstrong, Comparison of standard two-dimensional and three-dimensional corrected glenoid version measurements, J. Shoulder Elbow Surg. 20 (2011) 577–583, https://doi.org/10.1016/j.jse.2010.11.003.

[10] R.J. Friedman, K.B. Hawthorne, B.M. Genez, The use of computerized tomography in the measurement of glenoid version, J. Bone Joint Surg. 74 (1992) 1032–1037, https://doi.org/10.2106/00004623-199274070-00009.

[11] C.-H. Choi, H.-C. Kim, D. Kang, J.-Y. Kim, Comparative study of glenoid version and inclination using two-dimensional images from computed tomography and three-dimensional reconstructed bone models, Clin Shoulder Elbow 23 (2020) 119–124, https://doi.org/10.5397/cise.2020.00220.
[12] G.S. Lewis, A.D. Armstrong, Glenoid spherical orientation and version, J. Shoulder Elbow Surg. 20 (2011) 3–11, https://doi.org/10.1016/j.jse.2010.05.012.
[13] G. Moineau, C. Levigne, P. Boileau, A. Young, G. Walch, Three-dimensional measurement method of arthritic glenoid cavity morphology: Feasibility and reproducibility, Orthop. Traumatol. Surg. Res. 98 (2012) S139–S145, https://doi.org/10.1016/j.otsr.2012.06.007.
[14] A. Terrier, J. Ston, X. Larrea, A. Farron, Measurements of three-dimensional glenoid erosion when planning the prosthetic replacement of osteoarthritic shoulders, The Bone & Joint Journal 96-B (2014) 513–518, https://doi.org/10.1302/0301-620X.96B4.32641.
[15] P. Boileau, D. Cheval, M.-O. Gauci, N. Holzer, J. Chaoui, G. Walch, Automated three-dimensional measurement of glenoid version and inclination in arthritic shoulders, J. Bone Joint Surg. 100 (2018) 57–65, https://doi.org/10.2106/JBJS.16.01122.
[16] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat Methods 18 (2021) 203–211, https://doi.org/10.1038/s41592-020-01008-z.
[17] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, 3D Slicer as an image computing platform for the Quantitative Imaging Network, Magn. Reson. Imaging 30 (2012) 1323–1341, https://doi.org/10.1016/j.mri.2012.05.001.
[18] E. Taghizadeh, O. Truffer, F. Becce, S. Eminian, S. Gidoin, A. Terrier, A. Farron, P. Büchler, Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets, Eur Radiol 31 (2021) 181–190, https://doi.org/10.1007/s00330-020-07070-7.
[19] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, X.-J. Wu, Wing loss for robust facial landmark localisation with convolutional neural networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, USA, 2018, pp. 2235–2245.
[20] W. Schroeder, R. Maynard, B. Geveci, Flying edges: A high-performance scalable isocontouring algorithm, in: IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV), IEEE, Chicago, IL, USA, 2015, pp. 33–40, 10.1109/LDAV.2015.7348069.
[21] P.J. Denard, M.T. Provencher, A. Lädermann, A.A. Romeo, B.O. Parsons, J.S. Dines, Version and inclination obtained with 3-dimensional planning in total shoulder arthroplasty: do different programs produce the same results? JSES Open Access 2 (2018) 200–204, https://doi.org/10.1016/j.jses.2018.06.003.
[22] M. Minelli, A. Cina, F. Galbusera, A. Castagna, V. Savevski, L.M. Sconfienza, Measuring the critical shoulder angle on radiographs: an accurate and repeatable deep learning model, Skeletal Radiol 51 (2022) 1873–1878, https://doi.org/10.1007/s00256-022-04041-5.