



Automatic measurement of X-ray radiographic parameters based on cascaded HRNet model from the supraspinatus outlet radiographs

Yuwen Zheng^{1^}, Yuhua Wu^{2^}, Xiaofei Chen^{3^}, Ping Wang^{4^}, Fuwen Dong^{3^}, Linyang He^{5^}, Qing Su^{5^}, Guohua Cheng^{5^}, Chunyu Ma^{4^}, Hongyan Yao^{4^}, Sheng Zhou^{4^}

¹The First Clinical Medical College of Gansu University of Chinese Medicine, Lanzhou, China; ²Xi'an Hospital of Traditional Chinese Medicine, Xi'an, China; ³Department of Radiology, Gansu Provincial Hospital of Traditional Chinese Medicine, Lanzhou, China; ⁴Department of Radiology, Gansu Provincial Hospital, Lanzhou, China; ⁵Hangzhou Jianpei Technology Company Ltd., Hangzhou, China

Contributions: (I) Conception and design: Y Zheng; (II) Administrative support: G Cheng, S Zhou, P Wang; (III) Provision of study materials or patients: X Chen, F Dong; (IV) Collection and assembly of data: H Yao, C Ma; (V) Data analysis and interpretation: L He, Q Su, Y Zheng; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Sheng Zhou, MD. Department of Radiology, Gansu Provincial Hospital, No. 204 Donggang West Road, Chengguan District, Lanzhou 730000, China. Email: lzzs@sina.com.

Background: Rotator cuff injury is a common cause of shoulder pain. Precise and efficient measurement of morphological parameters is necessary in the clinical diagnosis and evaluation of shoulder disorders. However, manual measurement is a time-consuming and labor-intensive task, with low inter-observer reliability. The automatic measurement of radiographic parameters in supraspinatus outlet radiographs has not been reported yet. Thus, the objective of this study was to use a cascaded High-Resolution Net (HRNet) model based on deep learning (DL) algorithms to automatically measure morphological parameters from supraspinatus outlet radiographs and assess its performance. It was intended for use in early screening of patients with rotator cuff disease and to guide them to further consultation.

Methods: This cross-sectional study collected 1,668 supraspinatus outlet radiographs from the picture archiving and communication system of Gansu Provincial Hospital of Traditional Chinese Medicine and the Affiliated Hospital of Gansu University of Chinese Medicine. Among them, 521 images were provided for test datasets and 1,147 images were provided for a model training dataset and validation dataset. Landmarks were annotated for acromio-humeral interval (AHI), acromial tilt (AT), and 3 lines in Park's acromial classification (line huo-acrf, line acro-acro1, and line huo-acro1). R4 radiologist reviewed the means of 3 radiologists as a reference standard. Model performance was assessed by calculating the percentage of correct key points (PCK), intra-class correlation coefficients (ICCs), Pearson's correlation coefficients, mean absolute error, and root mean square error. The reliability of R1, R2, R3, AI with R4 and inter-observer reliability of R1, R2, and R3 for acromial morphology classification were assessed by Cohen's kappa coefficient.

Results: Within the 3-mm threshold, the PCK of the model ranged from 74% to 100%. Compared to the reference standard, the model had reliable measurement of AHI, AT, line huo-acrf, line acro-acro1, line huo-acro1 (ICC =0.73–0.94) and moderate reliability of acromial morphology classification ($k=0.50–0.56$).

Conclusions: The cascaded HRNet developed in this study can automatically measure morphological

[^] ORCID: Yuwen Zheng, 0000-0001-6286-9981; Yuhua Wu, 0000-0003-3054-4271; Xiaofei Chen, 0000-0003-2858-9817; Ping Wang, 0000-0002-9669-6246; Fuwen Dong, 0000-0003-4202-6957; Linyang He, 0000-0001-5181-8337; Qing Su, 0009-0006-1160-7969; Guohua Cheng, 0000-0003-4657-3691; Chunyu Ma, 0000-0003-2784-6826; Hongyan Yao, 0000-0002-9240-1535; Sheng Zhou, 0000-0002-9828-1968.

parameters of the shoulder. It may aid early clinical screening for shoulder disorders and assist physicians in treatment decisions.

Keywords: Deep learning (DL); shoulder; radiographic parameter; automatic measurement

Submitted Jul 06, 2024. Accepted for publication Dec 18, 2024. Published online Jan 22, 2025.

doi: 10.21037/qims-24-1373

View this article at: <https://dx.doi.org/10.21037/qims-24-1373>

Introduction

Lifetime prevalence of shoulder pain has been reported to be as high as 67% in the general population (1). Rotator cuff-related injuries are common causes of shoulder pain including subacromial impingement syndrome, rotator cuff tears, and subacromial pain syndrome, which may restrict activities of daily living, and contribute to loss of function and even disability (2,3). Imaging has played an important role in shoulder disorders, clinical decision, and prognostic assessment. X-rays are simple, have high spatial resolution, and are usually used as the preferred examination for shoulder pain. Measurement of morphological parameters from supraspinatus outlet radiographs, including acromio-humeral interval (AHI), acromial tilt (AT), and acromial morphology, has clinical significance for screening of rotator cuff disease, surgical options, and postoperative evaluation (4-6). Some studies have shown that supraspinatus outlet radiographs were superior to magnetic resonance imaging (MRI) for determination of acromial shapes and measurement of radiographic parameters (7,8). However, manual measurement is a time-consuming and labor-intensive task; in addition, there are some subjective errors of consistency and accuracy among different observers. Consequently, there is an urgent need for a new technology to solve the manual measurement challenge and improve disease detection rates.

Artificial intelligence (AI) has been computer-programmed to think and reason, and focuses on automatically solving intellectual tasks normally performed by the human brain (9). Deep learning (DL), as a branch of AI, can understand and imitate more complex processes due to its multi-hidden layers (10), and has gained great interest in the past few years, especially in the field of medical image analysis. It has laid a technical foundation for automatic measurement and has been initially explored in the field of shoulder, mainly focusing on shoulder image segmentation, disorders detection, and implant detection (11-13). The automatic measurement of radiographic parameters in

supraspinatus outlet radiographs has not yet been reported. The aim of this study was to construct a model based on the DL algorithm for fully automatic measurement of morphological radiographic parameters in supraspinatus outlet radiographs and to evaluate its performance. We present this article in accordance with the STROBE reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-1373/rc>).

Methods

Dataset preparation

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committees of Gansu Provincial Hospital of Traditional Chinese Medicine (No. 2024-046-01) and the Affiliated Hospital of Gansu University of Chinese Medicine [No. [2024] 101]. Due to the retrospective inclusion of imaging data from picture archiving and communication system (PACS), the requirement for patients' informed consent was waived by the respective ethics committees. From August 2017 to August 2023, a total of 2,107 supraspinatus outlet radiograph X-ray images and clinical data were collected in Gansu Provincial Hospital of Traditional Chinese Medicine; 234 supraspinatus outlet radiograph X-ray images and clinical data were collected in the Affiliated Hospital of Gansu University of Chinese Medicine. The inclusion criteria were as follows: (I) patients with complete epiphyseal closure; (II) cases in the positive test set were those with rotator cuff disease diagnosed by shoulder MRI. After assessment, 1,668 supraspinatus outlet radiograph X-ray images were finally included. The reasons of exclusion for 673 images were as follows: (I) irregularities in X-ray (n=458); (II) obscured landmarks and external body interference (n=187); (III) history of fractures, surgery, tumors, and tuberculosis (n=28). The data from Gansu Provincial Hospital of Traditional Chinese Medicine hospital were used for model training,

Table 1 Characteristics of patients in the training, validation, and test sets

Characteristic	Training set	Validation set	Internal test set	Positive test set	External test set
Total	899 [54]	248 [15]	215 [13]	148 [9]	158 [9]
Male	370 [41]	103 [42]	96 [45]	50 [34]	53 [33]
Female	529 [59]	145 [59]	119 [55]	98 [66]	105 [67]
Age (years)	53 [47, 60]	55 [48, 62]	52 [46, 58]	57 [50, 63]	56 [49, 63]
Male	52 [42, 58]	52 [47, 59]	51 [41, 58]	58 [49, 64]	57 [49, 63]
Female	54 [49, 61]	57 [50, 65]	53 [48, 59]	56 [50, 62]	56 [49, 63]

Data are expressed as numbers of patients with percentages or median, quartiles.

validation, an internal test set, and a positive test set, and the data from the Affiliated Hospital of Gansu University of Chinese Medicine Hospital were used for an external test set. A total of 215 (13%) cases were randomly assigned to the internal test set, 148 (9%) cases with rotator cuff disease diagnosed by shoulder MRI were assigned to the positive test set, and 158 (9%) cases were assigned to the external test set. The remaining 1,147 cases were randomly divided in a 3:1 ratio to form the training (n=899, 54%) and validation sets (n=248, 15%), respectively (*Table 1*).

Landmark annotation and parameter measurement

In this study, each radiograph has 8 landmarks named huo, hup, hup1, acro, acro1, acre, acrf, and cord (*Figure 1*). We measured 5 imaging parameters commonly used in supraspinatus outlet radiographs, including AHI, AT, and 3 lines of Park's acromial morphology (14), which were defined, respectively, as follows: AHI was the shortest distance between the inferior surface of the acromion and the head of the humerus; AT was angle between the line connecting the most anterior and last points of the subacromial surface and the line connecting the most anterior point of the subacromial surface and the lowest point of the coracoid; line huo-acrf; line acro-acro1; line huo-acro1. Type I acromion was defined as landmark acro coinciding with landmark acro1; type II acromion, the length of line huo-acrf, the same or longer than that of the line huo-acro1; type III acromion, the length of line huo-acrf, shorter than that of the line huo-acro (14) (*Figure 1*).

A total of four radiologists were involved in the data labelling and reviewing: 3 radiologists (R1, R2, and R3) labeled landmarks and R4 reviewed the finished labelled data using the JPHV-specific software (annotation 1.0.0; <http://47.110.61.145/>). A total of 1,147 radiograph were

used to train and validate the model, from which R1 manually annotated 9,176 landmarks, and R4 reviewed. All test sets were annotated independently by R1, R2, and R3, with the average of the 3 radiologists' measurements reviewed by R4 as the reference standard to analyze the consistency of the model with the reference standard and to assess the overall performance of the model. R1 re-labeled the internal test dataset after 4 weeks and analyzed the intra-observer reliability.

Data preprocessing

Data preprocessing adopted a 2-stage method. During the 2 stages of the model training phase, the data were augmented by random rotation (-10° to $+10^\circ$), random scaling (0.8 times), and horizontal flipping to increase the diversity of the data and improve the robustness of the model.

Model establishment and prediction

In this study, the cascaded High-Resolution Net (HRNet) model was trained to detect each key point, and the relevant parameters were calculated by coordinates of the key points. The HRNet used a 32-layer structure. The input channel input the images, which included 3 channels. The output of the model was 8 channels, with each channel outputting a heat map of 1 key point. The learning rate was initially set to 0.001 and decayed at 170th, 200th rounds to 0.0001. Optimization was performed with Adam optimizer, weight decay was set to $1e-5$, and loss function was conducted using MSELoss. During the model training phase, the original image was augmented, and then resized to 640×640 pixels for HRNet training in the first stage. Based on the coordinates of key points predicted by the first stage model, the minimum horizontal and vertical coordinates (min_x,

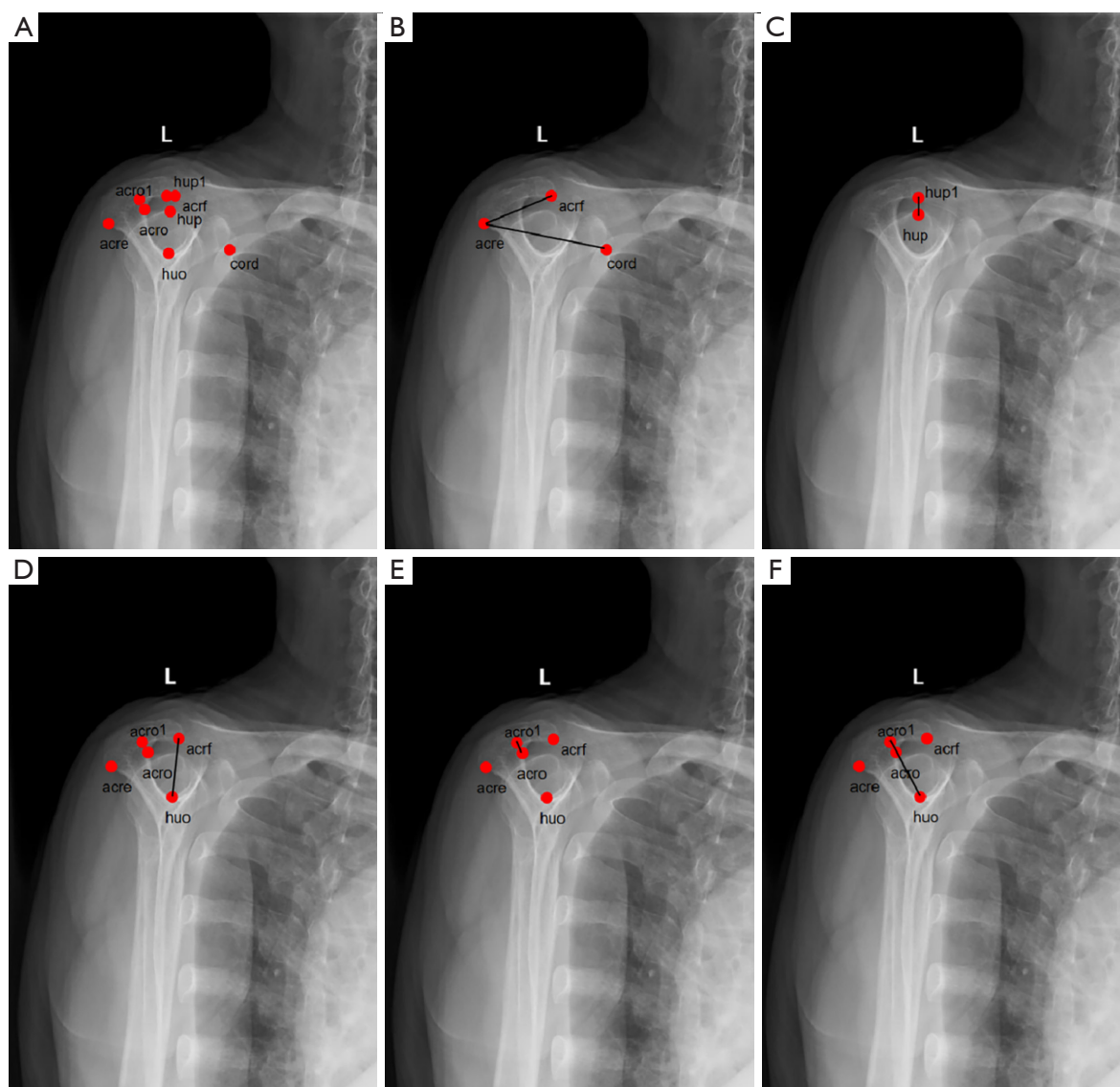


Figure 1 Schematic diagram of each landmark annotation and parameters measurement in supraspinatus outlet radiographs. (A) The specific name and anatomical location of each landmark. (B) Acromial tilt, the angle between the line acre-acrf and the line acre-cord. (C) Acromio-humeral interval, the distance between hup and hup1. (D) Line huo-acrf, the distance between huo and acrf. (E) Line acro-acro1, the distance between acro and acro1. (F) Line huo-acro1, the distance between huo and acro1. acre, the most posterior on the inferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; acro, the center of line acre-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion; huo, the center of the humeral head; hup, the apex of the humeral head; hup1, the point that the perpendicular line from hup with the cortex of the acromion; cord, the point on the lowermost edge of the coracoid.

min_y) were ascertained, as were the maximum horizontal and vertical coordinates (max_x, max_y). The region of interest (ROI; min_x-200: max_x+200, min_y-200: max_y+200) was cut out from the original image. The ROI was also augmented, and then resized to 640×640 pixels for the

second-stage model training.

During the model inference phase, the whole image was resized to 640×640 pixels as the input for the first stage model to detect each key point, and 8 heatmaps with 160×160 pixels were output. The coordinates with the

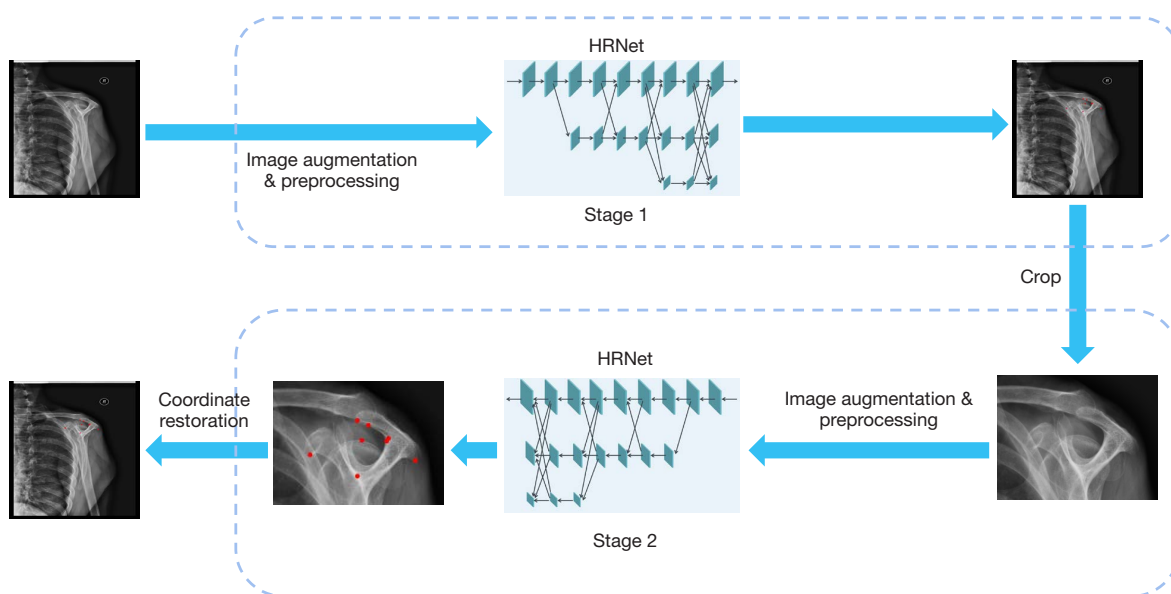


Figure 2 Model prediction process. HRNet, High-Resolution Net.

highest values were taken from each heatmap and resized to the original size to get the coordinates. Based on the coordinates of the key points predicted in the first stage, the ROI was cut out and resized to 640×640 pixels for the second stage model prediction to get the coordinates, and then resized to the original image size to obtain the final coordinates (*Figure 2*).

Statistical analysis

The data were analyzed using Python (scipy.stats, bland-altman, sklearn.metrics.regression, etc.) (Python Software Foundation, Wilmington, DE, USA) Microsoft Excel 2021 (Microsoft Corp., Redmond, WA, USA), and SPSS 26.0 (IBM Corp., Armonk, NY, USA). The differences were considered statistically significant with a P value <0.05. If the data followed normal distribution and homogeneity of variance, the paired *t*-test was used for comparison between groups, otherwise the rank-sum test was used. The distribution of information such as age and gender of the patients in the training, validation, and test sets were statistically described using medians and percentages.

Reliability of landmark annotations

The percentage of correct key points (PCK) within the 1-, 2-, 3-, 4-, and 5-mm key point-to-key point distance thresholds were calculated to assess intra- and inter-

observer image annotation consistency.

Landmark performance

The PCK was used to access the performance of the model in predicting all landmarks. PCK (15) was defined as the proportion of computationally detected key points for which the corresponding normalized distance between reference standards was less than a set threshold value.

Comparison of convolutional neural networks

To further compare the accuracy of the models in predicting key points, the cascaded HRNet was compared with HRNet (16), self-calibrated convolutions (SCNet) (17), and U-Net (18). For objective and fair comparisons, the same data training and strategies were used for all models.

Model measurement performance

This study used the average of the 3 radiologists' measurements reviewed by R4 as the reference standard, and calculated measurements from R1, R2, R3, R4, and model predictions. We compared the model with the reference standard and each radiologist separately. To evaluate the overall performance of the cascade HRNet, the intra-class correlation coefficient (ICC), Pearson correlation

Table 2 The intra- and inter-observer reliability of landmark annotation (%)

Radiologist	Threshold (mm)				
	1	2	3	4	5
Inter-observer	63	84	92	96	98
Intra-observer					
R1 vs. R2	70	90	96	98	99
R2 vs. R3	47	72	85	92	96
R3 vs. R1	51	74	87	93	96

coefficient (r), mean absolute error (MAE), and root mean square error (RMSE) were compared between the reference standard and model predictions from test sets. ICC was used to measure and evaluate the consistency between the reference standard and model predictions, with $\text{ICC} \geq 0.75$ indicating excellent reliability. $|r| \geq 0.7$ indicated high correlation. MAE was defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\text{observed}_i - \text{predicted}_i| \quad [1]$$

and RMSE was defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{observed}_i - \text{predicted}_i)^2} \quad [2]$$

where i denotes reference and predicted values for the i th image, and n is the number of images.

In addition, Bland-Altman plots showed the mean difference, standard deviation (SD), and 95% limit of agreement (95% LoA) between the reference standard and the model measurements. In order to compare the performance of the model with each radiologist, the mean of each of these 2 radiologists was made a paired difference with the model and the third radiologist's measurements, respectively. Paired t -tests were performed on the paired differences between the 2 groups in order to compare these MAEs and determine statistical differences.

Reliability analysis of acromial morphology classification

According to the Park's criteria, this study calculated acromial morphology in the internal test set and positive test set based on the line acro-acro1, huo-acrf, and huo-acro1 values measured by R1, R2, R3, R4, and the model. Kappa test was used to analyze the acromial morphology consistency of AI with R4 and inter-observer consistency

of R1, R2, and R3. The differences were considered statistically significant when $P < 0.05$.

Results

Reliability of landmark annotation

Intra- and inter-observer consistency analysis showed that the intra- and inter-observer PCK were greater than 85% within the 3 mm distance threshold (Table 2).

Landmark performance

The PCK at the 3 mm distance threshold were 74–100%, with the largest PCK for acro and the smallest PCK for cord (Table 3, Figure 3). Heat maps of model prediction are shown in Figure 4.

Comparison of CNNs

The PCK of our model within the 3 mm threshold range was greater than that of HRNet, SCNet, and U-Net, especially between the 1 and 2 mm threshold ranges (Table 4, Figure 5).

Model measurement performance

The results of paired t -tests of the model predictions and reference standard showed that in all test sets, the AT and acro-acro1 distance were not significantly different from the reference standard ($P > 0.05$), AHI, huo-acrf, huo-acro1 showed statistically significant differences ($P < 0.05$) (Table 5).

To further assess the overall performance of the model, the model showed an excellent ICC ranging from 0.73 to 0.94, high correlation (r) ranging from 0.73 to 0.95, MAE ranging from 0.58 to 1.70, and RMSE ranging from 0.78 to 2.16 (Table 5). Correlation scatter plots and Bland-Altman plots showed the difference between model and reference standard (Figure 6).

In addition, we compared the model's differences with each radiologist. For AHI and AT, the MAEs predicted by the model were significantly less than the MAEs of 3 radiologists ($P < 0.05$). For line acro-acro1, the MAEs of the model were significantly less than R3 ($P < 0.05$) and comparable in performance to R1 ($P = 0.18$) and R2 ($P = 0.05$). For line huo-acrf, the MAEs of the model were significantly less than R1 and R2 ($P < 0.001$). For line huo-acro1, the MAEs of the model were significantly higher than R1

Table 3 The PCK values of landmarks at the 1–5 mm threshold (%)

Key point	Threshold (mm)				
	1	2	3	4	5
L					
acre	51	78	89	96	98
acrf	64	90	95	97	97
acro	60	92	100	100	100
acro1	55	87	97	99	100
cord	39	62	74	86	92
huo	23	66	88	95	97
hup	54	80	91	97	100
hup1	29	60	76	87	92
R					
acre	53	84	92	98	98
acrf	63	89	95	97	98
acro	61	95	99	100	100
acro1	56	83	96	98	99
cord	39	67	75	88	90
huo	21	70	89	96	96
hup	53	77	91	96	100
hup1	33	65	82	88	92

acre, the most posterior on the inferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; acro, the center of line acre-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion, hu0, the center of the humeral head; hup, the apex of the humeral head; hup1, the point that the perpendicular line from hup with the cortex of the acromion; cord, the point on the lowermost edge of the coracoid. PCK, percentage of correct key points; L, left shoulder; R, right shoulder.

($P < 0.001$), less than R2 ($P < 0.001$), and with R3 were not statistically significant (Table 6).

Reliability of acromial morphology classification

The standard reference of acromial morphology was calculated by R4's measurements. In the internal test set, the agreement rating of R1 and R2 were 83% and 87%, respectively. R1 and R2 had good reliability with R4, and kappa values of 0.66 and 0.74, respectively. The agreement ratings of R3 and model were 67% and 75%, respectively. R3 and model had moderate reliability with R4, and kappa

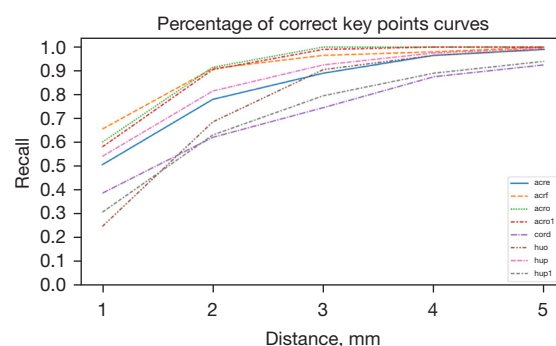


Figure 3 The ability of cascaded HRNet to predict each landmark. acre, the most posterior on the inferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; acro, the center of line acre-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion, hu0, the center of the humeral head; hup, the apex of the humeral head; hup1, the point that the perpendicular line from hup with the cortex of the acromion; cord, the point on the lowermost edge of the coracoid. HRNet, High-Resolution Net.

values of 0.42 and 0.50, respectively (Table 7). R3 had general reliability with R1 and R2, with kappa values of 0.38 and 0.32, respectively. R1 had good reliability with R2, with a kappa value of 0.63 (Tables 7,8).

In the positive test set, the agreement rating of R1, R2, and R3 were above 80%, all with good reliability and kappa values of 0.76, 0.74, and 0.68, respectively (Table 7). The agreement rating of the model was 80%. The model had moderate reliability with R4, and a kappa value of 0.56 (Table 7). R1, R2, and R3 inter-observer kappa values were also all above 0.50 (Table 8).

Discussion

There are few studies on DL in the field of automatic measurement in shoulder X-ray radiography. We explored methods to automatically measure supraspinatus outlet radiographs parameters based on DL. The cascaded HRNet model performed better than other models, automatically identified landmarks, and measured parameters more accurately. In addition, the cascaded HRNet could accurately measure AHI, AT, hu0-acro1, acro-acro1, and hu0-acrf (ICC =0.73–0.94). Minelli *et al.* (19) measured the critical shoulder angle (CSA) on radiographs based on Inception V3 coupled with a spatial to numerical transform layer, and the model predicted CSA at a median error of 0.95° with a standard deviation of 0.97°. Furthermore, some

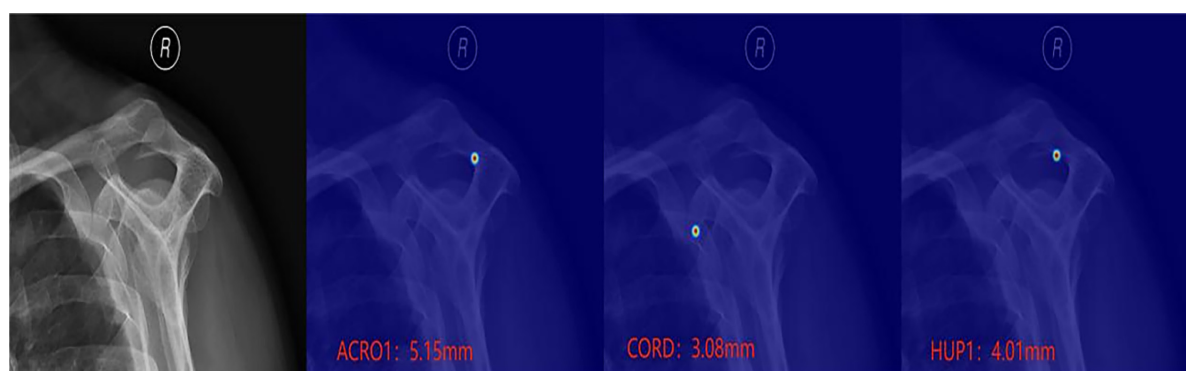


Figure 4 Cascaded HRNet model prediction heatmap. acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion; cord, the point on the lowermost edge of the coracoid; hup1, the point that the perpendicular line from hup with the cortex of the acromion. HRNet, High-Resolution Net.

Table 4 Comparison of prediction of landmarks by cascaded HRNet, HRNet, SCNet, U-Net (%)

Model	Threshold (mm)				
	1	2	3	4	5
HRNet	33	71	86	93	96
SCNet	28	65	84	93	97
U-Net	37	71	86	94	96
Cascaded HRNet	47	75	86	93	96

HRNet, High-Resolution Net; SCNet, self-calibrated convolutions.

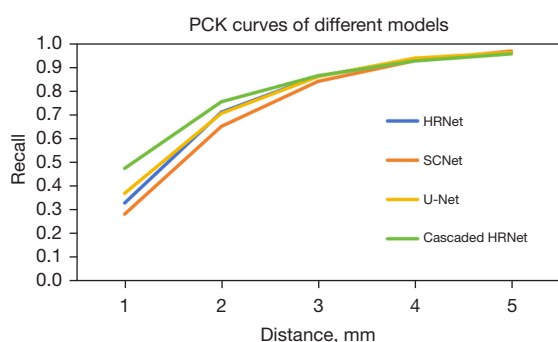


Figure 5 The comparison of the landmark prediction by U-Net, HRNet, cascaded HRNet, SCNet. PCK, percentage of correct key points; HRNet, High-Resolution Net; SCNet, self-calibrated convolutions.

studies have also used DL techniques to diagnose relevant diseases in radiographic images with computed tomography (CT) or MRI as the gold standard. For example, some researchers developed DL models to diagnose COVID-19

disease symptoms on chest X-rays and achieved high sensitivity and specificity (20). Chen *et al.* (21) used MRI as the reference standard and showed that DL models could identify fresh vertebral compression fractures from lumbar X-rays. Most of these DL algorithms were for automatic classification, which lacked precise quantitative assessment. The cascaded HRNet could provide actual values for surgeons to perform more objective assessment. If the 2 kinds of models were combined, automatic diagnosis of diseases as well as numerical assessment of severity may be achieved, and this represents a topic for future exploration.

The great performance of our model is relevant with image quality control, pre-processing, data set allocation, and model construction. In order to ensure the quality of the training data and to accurately label landmarks, this study strictly followed the quality control criteria. In the data pre-processing, the augmentation approach in this study could make the best use of limited data and increase the number and diversity of images (22). In the data set allocation, the training and validation sets contain multiple diseases, including healthy individuals, rotator cuff tendon tear, subacromial impingement syndrome, and so on, as well as a wide age range. Thus, the model has better application for complex clinical situations with higher robustness (23). In addition, data from other hospitals were collected to assess the model's generalization ability, and the model predictions also had good performance (ICC =0.73–0.93). Contrastingly, Kufel *et al.* (24) used natural language processing to extract disease classifications via text from relevant radiology reports. This method enabled the collection of general disease and reduced workload while improving efficiency, which is an advantageous method for collecting diseases in the future.

Table 5 The comparison of model and the standard reference

Parameter	R4	Model	<i>t</i>	P value	ICC	<i>r</i>	MAE	RMSE
Internal test set								
AHI (mm)	8.79±1.88	9.18±1.87	2.10	0.04*	0.92	0.94	0.58	0.78
AT (°)	37.04±5.57	36.56±5.78	−0.83	0.41	0.94	0.94	1.53	2.03
acro-acro1 (mm)	3.98±1.70	3.79±1.49	−1.17	0.25	0.83	0.84	0.72	0.94
huo-acrf (mm)	30.86±3.05	31.92±2.91	3.56	<0.001*	0.83	0.88	1.38	1.81
huo-acro1 (mm)	29.01±2.94	30.02±2.64	3.62	<0.001*	0.82	0.88	1.35	1.74
Positive test set								
AHI (mm)	8.66±1.78	9.08±1.72	1.97	0.05*	0.91	0.93	0.59	0.78
AT (°)	36.85±5.62	36.55±5.34	−0.45	0.65	0.94	0.95	1.41	1.84
acro-acro1 (mm)	4.07±1.65	4.08±1.41	0.02	0.99	0.79	0.80	0.74	1.00
huo-acrf (mm)	30.27±2.53	31.19±2.40	3.08	0.002*	0.78	0.84	1.30	1.69
huo-acro1 (mm)	28.89±2.27	29.91±2.18	3.75	<0.001*	0.75	0.83	1.28	1.65
External test set								
AHI (mm)	8.52±1.93	9.07±1.81	2.49	0.01*	0.87	0.91	0.75	0.98
AT (°)	38.22±5.61	37.93±5.50	−0.44	0.66	0.93	0.93	1.68	2.11
acro-acro1 (mm)	3.92±1.43	3.96±1.26	0.22	0.83	0.73	0.73	0.77	1.00
huo-acrf (mm)	30.77±2.86	32.26±2.88	4.38	<0.001*	0.78	0.88	1.61	2.06
huo-acro1 (mm)	28.52±3.14	30.05±2.81	4.35	<0.001*	0.77	0.88	1.70	2.16

R4 is the reference standard. Model is the parameter measured by the model. The values are presented as means ± SD. *, $P_{(\text{paired } t\text{-test})} < 0.05$ indicates statistical significance between the model and the reference standard. acro, the center of line acro-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; huo, the center of the humeral head. ICC, intra-class correlation coefficient; *r*, Pearson correlation coefficient; MAE, mean absolute error; RMSE, root mean square error; AHI, acromio-humeral interval; AT, acromial tilt; acro-acro1, line acro-acro1; huo-acrf, line huo-acrf; huo-acro1, line huo-acro1.

In terms of model selection, U-Net, SCNet, HRNet, and cascaded HRNet were selected for training on the same portion of data, and the above models showed great performance in other tasks (25–27). The best performing cascaded HRNet model was selected based on model predictions. Cascaded HRNet is based on HRNet and uses a 2-stage approach to predict key points. HRNet uses multi-resolution subnetworks to maintain high resolution throughout the process (16). Considering the overlapping of skeletal images, a cascaded design could gradually improve the precision of key points prediction.

This study set an independent positive test set, which could verify whether the model could accurately measure the imaging parameters of rotator cuff tear. The results suggested that the model predictions were in good agreement with the reference standard (ICC: 0.75–0.94) but were slightly lower than those in the internal test set. The reason may be that

there were less data of severe rotator cuff tear in the training set, which led to the model's insufficient learning for these images. The mean of AT in the positive test set was 36.85°, which was less than 37° and was consistent with the findings of previous studies (6,28,29). However, the mean of AHI was 8.66 mm, which was higher than the cutoff of 7 mm in previous studies, but less than 10 mm. The narrowing of the subacromial space in some patients with rotator cuff tears are not significant. Therefore, we should be wary of rotator cuff tears or subacromial impingement when AHI is less than 10 mm (30,31).

Accuracy of landmark prediction is the main cause of parameter measurement errors. In the present study, inter-observer PCK within the 3 mm threshold was above 84%, indicating that our 3 radiologists were relatively reliable. The model was able to predict most of the key points. However, cord, hup1, and huo were relatively poorly

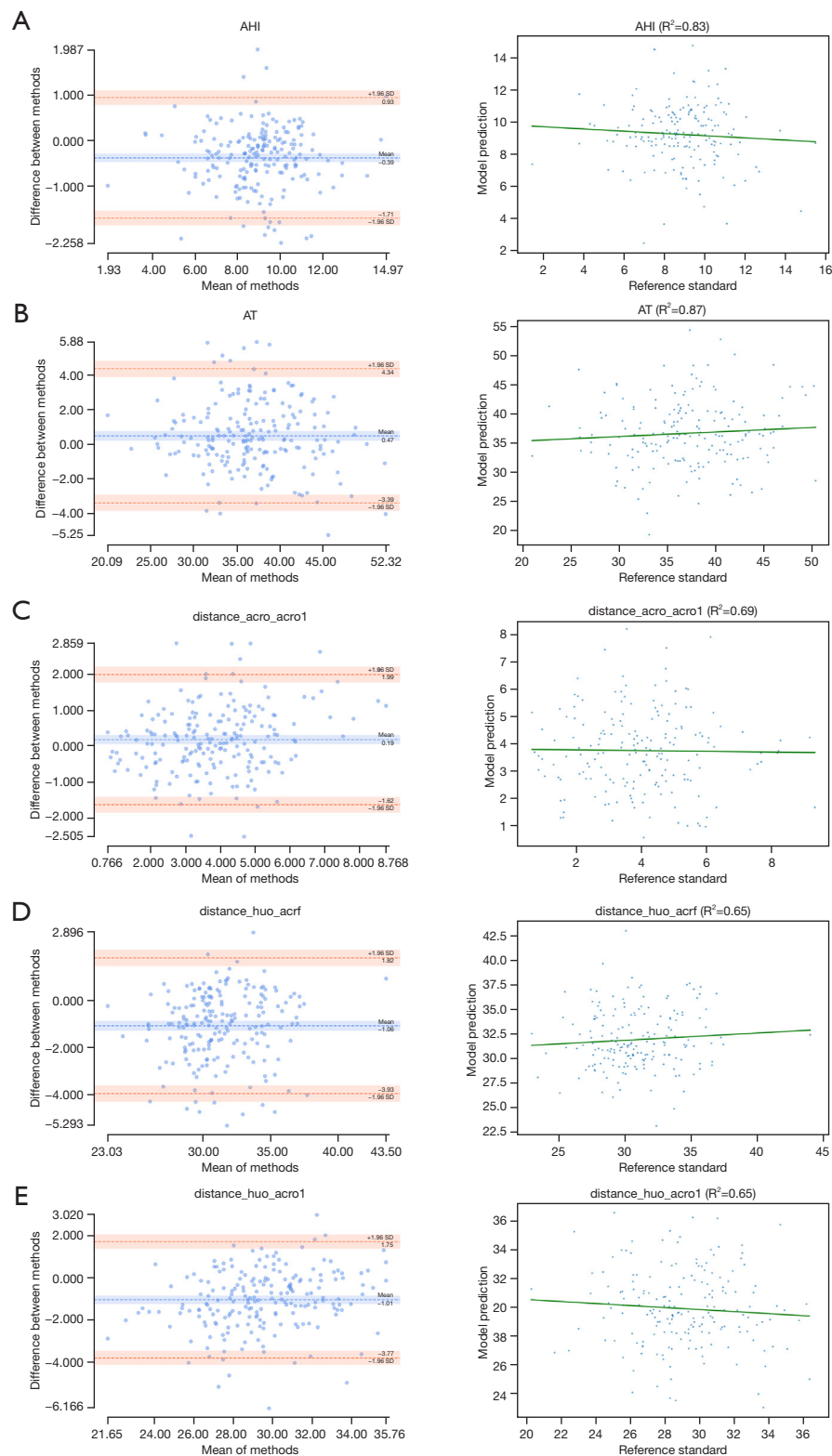


Figure 6 Comparison of our model prediction with the reference standard, the Bland-Altman plots on the left, the correlation scatter plots on the right. (A) AHI; (B) AT; (C) line acro-acro1; (D) line huo-acrf; (E) line huo-acro1. acro, the center of line acre-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; huo, the center of the humeral head. AHI, acromio-humeral interval; AT, acromial tilt.

Table 6 Comparisons of each radiologist and the model for supraspinatus outlet parameters

Parameter	Mean of R2 and R3	Mean of R1 and R3	Mean of R1 and R2
AHI (mm)			
R*	2.07	2.11	2.05
Model	2.07	0.57	0.58
P value	0.03*	0.001*	0.002*
AT (°)			
R*	6.70	6.54	6.74
Model	6.50	1.48	1.71
P value	0.02*	0.03*	0.001*
acro-acro1 (mm)			
R*	1.94	1.96	1.92
Model	1.81	0.76	0.82
P value	0.18	0.05	<0.001*
huo-acrf (mm)			
R*	3.26	3.17	3.53
Model	3.21	1.03	1.46
P value	<0.001*	<0.001*	0.25
huo-acro1 (mm)			
R*	3.02	3.07	3.28
Model	3.06	1.12	1.37
P value	<0.001*	<0.001*	0.34

, P<0.05 indicates a statistically significant inter-observer difference. R represents the third radiologist. acro, the center of line acro-acrf; acro1, the point that the perpendicular line from acro with the anteroinferior cortex of the acromion; acrf, the most anterior on the inferior cortex of the acromion; huo, the center of the humeral head. AHI, acromio-humeral interval; AT, acromial tilt; acro-acro1, line acro-acro1; huo-acrf, line huo-acrf; huo-acro1, line huo-acro1.

Table 7 The reliability of R1, R2, R3, AI with R4 on the internal and positive test set

Data sets	R4 vs. R1		R4 vs. R2		R4 vs. R3		R4 vs. model	
	Agreement	k	Agreement	k	Agreement	k	Agreement	k
Internal test set	83%	0.66	87%	0.74	67%	0.42	75%	0.50
Positive test set	88%	0.76	87%	0.74	84%	0.68	80%	0.56

Table 8 The inter-observer reliability of acromial morphology classification on the internal and positive test set

Data sets	R1 vs. R2	R1 vs. R3	R2 vs. R3
Internal test set	0.63	0.38	0.32
Positive test set	0.72	0.56	0.52

predicted. The reasons may be as follows: (I) for huo, it was obscure because of the overlapping relationship between scapula and humeral head. In addition, it was stated in the center of humeral head instead of clear marker of bone. (II) Acromial morphology varied widely in some individuals to lead great variation for hup1. Secondly, the acromion and the humeral head are subcurved in the type II acromion,

which means that each distance between humerus head and inferior surface of the acromion are similar. The model fails to identify the corresponding hup and hup1, which may lead larger AHI. (III) The cord may overlap with outer edge of the ribs, which increases challenge for identification.

In all test sets, measurement of the model was consistent and relevant with the standard reference (ICC =0.73–0.94, $r=0.73-0.95$). However, the results of paired *t*-tests showed that AHI, huo-acrf, and huo-acro1 of the model were significant with the standard reference ($P<0.05$), and Bland-Altman plots indicated that several data were exceeding the 95% limit of agreement. It may be related with the moderate accuracy of hup1 and huo's predictions. In addition, the distance between hup and hup1 in patients with a severely narrow subacromial space was closer, which may increase the demand for measurement precision. Thus, even slight landmarks errors impact on parameter measurement. On further analysis of the model with each radiologist's variance, it was found that most of the model's MAEs were significantly smaller than those of the radiologists, suggesting that the model performed close to the radiologists. Kufel *et al.* (32) used transfer learning techniques to automatically detect chest X-ray abnormalities. The advantage of transfer learning techniques is that it could improve accuracy for model application in small sample disease. In the next stage of optimizing model performance, it could be used to improve the model's accuracy of the automatic measurement in rotator cuff disease data with significant narrowing AHI.

The most common acromial morphology classification was proposed by Bigliani *et al.*, describing a flat (type-I), curved (type-II), or hooked (type-III) morphology. It is a qualitative assessment with subjective errors among observers, especially for type II and III acromion (7). In this study, we used Park's criteria, which is able to classify the acromion more accurately by quantifying the inter-anatomical distances (14). It is more suitable for DL to classify acromial morphology. According to Park's approach for type I, it is impossible that acro coincide completely with acro1 because of a slight error in manual annotation. Chen *et al.* (33) reported that a landmark-to-landmark distance of 2.98 mm for inter-observer observations was acceptable. Thus, we defined that line acro-acro1 being shorter or equal to 1.5 mm constituted as type-I, and only that longer than 1.5 mm would be classified type II or type III. The poor interobserver agreement of Bigliani's criteria was reported in a previous study (34). In the present study, R1 and R2 had good reliability with R4, with the kappa values all above 0.6 in the positive and internal test sets. The reliability of

R3 with R1, R2, and R4 was poor to moderate, and the kappa value of model was higher than that of R3, indicating that the model reduced subjective error to a certain extent.

Some data were inconsistent for type II and III. In some images, partial overlap of the clavicle with the acromion leads to difficulty in locating acrf, affecting the measurement of the line huo-acrf. In addition, several data showed a hierarchical leap. For example, when 3 radiologists classified type I, another classified type II or type III. This may have been caused by inaccurate measurements of line acro-acro1 due to the mistake of acro1. During the labelling process, we found that the location of acro was different in 3 cases, including on the edge of cortical bone, outside the bone, and on the bone. For the first and second cases, the location of acro1 are better determined. When the acro is located on the bone, there may be a divergence in the labelling of acro1 on the undersurface of the acromion, which can lead to a large difference in acromial morphology.

This study has some limitations. Firstly, the training datasets only included 899 images, which is relatively small for training a model. Secondly, some difference remained in the external dataset compared to internal dataset. Future studies will be planned to incorporate more disease types, expand the sample size, and improve the accuracy and applicability of the model to empower clinical diagnosis and treatment. Secondly, we will improve Park's approach according to reality and develop a suitable labelling method for our requirements. Finally, we will analyze the association between diseases and parameters. The model will gradually apply to the clinical environment to provide structured reports of quantitative parameters.

Conclusions

The impressive results obtained by our model suggest its usefulness in the automatic measurement of morphological parameters from supraspinatus outlet radiographs, including classification of acromial morphology, which could help early clinical screening of rotator cuff disease and assist doctors in treatment decision-making.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at <https://qims.amepub.com/qims>.

amegroups.com/article/view/10.21037/qims-24-1373/rc

Funding: This study was supported by the Natural Science Foundation of Gansu Province, China (Nos. 22JR5RA699 and 22JR5RA659) and the National Natural Science Foundation, China (No. 8236070228).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-1373/coif>). G.C. is a consultant of Hangzhou Jianpei Technology Co., Ltd. L.H. and Q.S. are employees of Hangzhou Jianpei Technology Co., Ltd. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committees of Gansu Provincial Hospital of Traditional Chinese Medicine (No. 2024-046-01) and the Affiliated Hospital of Gansu University of Chinese Medicine [No. [2024] 101]. Due to the retrospective inclusion of imaging data from picture archiving and communication system (PACS), the requirement for patients' informed consent was waived by the ethics committees.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, Verhaar JA. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33:73-81.
2. Lafrance S, Charron M, Dubé MO, Desmeules F, Roy JS, Juul-Kristensen B, Kennedy L, McCreesh K. The Efficacy of Exercise Therapy for Rotator Cuff-Related Shoulder Pain According to the FITT Principle: A Systematic Review With Meta-analyses. *J Orthop Sports Phys Ther* 2024;54:499-512.
3. Sachinis NP, Yiannakopoulos CK, Chalidis B, Kitridis D, Givissis P. Biomolecules Related to Rotator Cuff Pain: A Scoping Review. *Biomolecules* 2022;12:1016.
4. Saupé N, Pfirrmann CW, Schmid MR, Jost B, Werner CM, Zanetti M. Association between rotator cuff abnormalities and reduced acromiohumeral distance. *AJR Am J Roentgenol* 2006;187:376-82.
5. Yang J, Xiang M, Li Y, Zhang Q, Dai F. The Correlation between Various Shoulder Anatomical Indices on X-Ray and Subacromial Impingement and Morphology of Rotator Cuff Tears. *Orthop Surg* 2023;15:1997-2006.
6. Andrea LC, Svendsen SW, Frost P, Smidt K, Gelineck J, Christiansen DH, Deutch SR, Hansen TB, Haahr JP, Dalbøge A. Radiographic findings in patients suspected of subacromial impingement syndrome: prevalence and reliability. *Skeletal Radiol* 2024;53:2477-90.
7. Sahin K, Kendirci AS, Kocazeybek E, Demir N, Sağlam Y, Ersen A. Reliability of Bigliani's Classification using Magnetic Resonance Imaging for Determination of Acromial Morphology. *Malays Orthop J* 2022;16:44-9.
8. Suter T, Krähenbühl N, Howell CK, Zhang Y, Henninger HB. Viewing perspective malrotation influences angular measurements on lateral radiographs of the scapula. *J Shoulder Elbow Surg* 2020;29:1030-9.
9. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* 2017;37:505-15.
10. Kufel J, Bargiel-Łączek K, Kocot S, Koźlik M, Bartnikowska W, Janik M, Czogalik Ł, Dudek P, Magiera M, Lis A, Paszkiewicz I, Nawrat Z, Cebula M, Gruszczyńska K. What Is Machine Learning, Artificial Neural Networks and Deep Learning?-Examples of Practical Applications in Medicine. *Diagnostics (Basel)* 2023;13:2582.
11. Taghizadeh E, Truffer O, Becce F, Eminian S, Gidoin S, Terrier A, Farron A, Büchler P. Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets. *Eur Radiol* 2021;31:181-90.
12. Yao J, Chepelev L, Nisha Y, Sathiadoss P, Rybicki FJ, Sheikh AM. Evaluation of a deep learning method for the automated detection of supraspinatus tears on MRI. *Skeletal Radiol* 2022;51:1765-75.
13. Yi PH, Kim TK, Wei J, Li X, Hager GD, Sair HI, Fritz J. Automated detection and classification of shoulder arthroplasty models using deep learning. *Skeletal Radiol* 2020;49:1623-32.

14. Park TS, Park DW, Kim SI, Kweon TH. Roentgenographic assessment of acromial morphology using supraspinatus outlet radiographs. *Arthroscopy* 2001;17:496-501.
15. Payer C, Štern D, Bischof H, Urschler M. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Med Image Anal* 2019;54:207-19.
16. Sun K, Xiao B, Liu D, Wang J, editors. Deep High-Resolution Representation Learning for Human Pose Estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 15-20 June 2019.
17. Liu JJ, Hou Q, Cheng MM, Wang C, Feng J, editors. Improving Convolutional Networks With Self-Calibrated Convolutions. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 13-19 June 2020.
18. Ronneberger O, Fischer P, Brox T, editors. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015 2015//; Cham: Springer International Publishing.
19. Minelli M, Cina A, Galbusera F, Castagna A, Savevski V, Sconfienza LM. Measuring the critical shoulder angle on radiographs: an accurate and repeatable deep learning model. *Skeletal Radiol* 2022;51:1873-8.
20. Kufel J, Bargieł K, Koźlik M, Czogalik Ł, Dudek P, Jaworski A, Cebula M, Gruszczyńska K. Application of artificial intelligence in diagnosing COVID-19 disease symptoms on chest X-rays: A systematic review. *Int J Med Sci* 2022;19:1743-52.
21. Chen W, Liu X, Li K, Luo Y, Bai S, Wu J, Chen W, Dong M, Guo D. A deep-learning model for identifying fresh vertebral compression fractures on digital radiography. *Eur Radiol* 2022;32:1496-505.
22. Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: A systematic literature review. *Comput Biol Med* 2023;152:106391.
23. Conze PH, Brochard S, Burdin V, Sheehan FT, Pons C. Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders. *Comput Med Imaging Graph* 2020;83:101733.
24. Kufel J, Bargieł-Łączek K, Koźlik M, Czogalik Ł, Dudek P, Magiera M, Bartnikowska W, Lis A, Paszkiewicz I, Kocot S, Cebula M, Gruszczyńska K, Nawrat Z. Chest X-ray Foreign Objects Detection Using Artificial Intelligence. *J Clin Med* 2023;12:5841.
25. Wu Y, Chen X, Dong F, He L, Cheng G, Zheng Y, Ma C, Yao H, Zhou S. Performance evaluation of a deep learning-based cascaded HRNet model for automatic measurement of X-ray imaging parameters of lumbar sagittal curvature. *Eur Spine J* 2024;33:4104-18.
26. Yan Y, Zhang X, Meng Y, Shen Q, He L, Cheng G, Gong X. Sagittal intervertebral rotational motion: a deep learning-based measurement on flexion-neutral-extension cervical lateral radiographs. *BMC Musculoskelet Disord* 2022;23:967.
27. Trinh GM, Shao HC, Hsieh KL, Lee CY, Liu HW, Lai CW, Chou SY, Tsai PI, Chen KJ, Chang FC, Wu MH, Huang TJ. Detection of Lumbar Spondylolisthesis from X-ray Images Using Deep Learning Network. *J Clin Med* 2022;11:5450.
28. Zuckerman JD, Kummer FJ, Cuomo F, Simon J, Rosenblum S, Katz N. The influence of coracoacromial arch anatomy on rotator cuff tears. *J Shoulder Elbow Surg* 1992;1:4-14.
29. Aoki M, Ishii S, Usui M, Mizuguchi M, Miyano S. The slope of the acromion and rotator cuff impingement. *Katakansetsu* 1986;10:168-71.
30. Nové-Josserand L, Lévine C, Noël E, Walch G. The acromio-humeral interval. A study of the factors influencing its height. *Rev Chir Orthop Reparatrice Appar Mot* 1996;82:379-85.
31. Jeong JH, Yoon EJ, Kim BS, Ji JH. Biceps-incorporating rotator cuff repair with footprint medialization in large-to-massive rotator cuff tears. *Knee Surg Sports Traumatol Arthrosc* 2022;30:2113-22.
32. Kufel J, Bielówka M, Rojek M, Mitreğa A, Lewandowski P, Cebula M, Krawczyk D, Bielówka M, Kondoł D, Bargieł-Łączek K, Paszkiewicz I, Czogalik Ł, Kaczyńska D, Wocław A, Gruszczyńska K, Nawrat Z. Multi-Label Classification of Chest X-ray Abnormalities Using Transfer Learning Techniques. *J Pers Med* 2023;13:1426.
33. Chen HC, Lin CJ, Wu CH, Wang CK, Sun YN. Automatic Insall-Salvati ratio measurement on lateral knee x-ray images using model-guided landmark localization. *Phys Med Biol* 2010;55:6785-800.
34. McLean A, Taylor F. Classifications in Brief: Bigliani Classification of Acromial Morphology. *Clin Orthop Relat Res* 2019;477:1958-61.

Cite this article as: Zheng Y, Wu Y, Chen X, Wang P, Dong F, He L, Su Q, Cheng G, Ma C, Yao H, Zhou S. Automatic measurement of X-ray radiographic parameters based on cascaded HRNet model from the supraspinatus outlet radiographs. *Quant Imaging Med Surg* 2025;15(2):1425-1438. doi: 10.21037/qims-24-1373