



ELSEVIER

Contents lists available at ScienceDirect

JSES Reviews, Reports, and Techniques

journal homepage: www.jsesreviewsreportstech.org

Deep learning model for measurement of shoulder critical angle and acromion index on shoulder radiographs



M. Moein Shariatnia, MD^a, Taghi Ramazanian, MD^{b,c}, Joaquin Sanchez-Sotelo, MD, PhD^c, Hilal Maradit Kremers, MD, MSc^{b,c,*}

^aMedical Student, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

^bDepartment of Quantitative Sciences Research, Mayo Clinic, Rochester, MN, USA

^cDepartment of Orthopedic Surgery, Rochester, MN, USA

ARTICLE INFO

Keywords:

Critical shoulder angle
Acromion index
Deep learning
Rotator cuff tear
Glenohumeral osteoarthritis
Artificial intelligence

Level of evidence: Level III; Retrospective Study

Background: Several bone morphological parameters, including the anterior acromion morphology, the lateral acromial angle, the coracohumeral interval, the glenoid inclination, the acromion index (AI), and the shoulder critical angle (CSA), have been proposed to impact the development of rotator cuff tears and glenohumeral osteoarthritis. This study aimed to develop a deep learning tool to automate the measurement of CSA and AI on anteroposterior shoulder radiographs.

Methods: We used MURA Dataset v1.1, which is a large publicly available musculoskeletal radiograph dataset from the Stanford University School of Medicine. All normal shoulder anteroposterior radiographs were extracted and annotated by an experienced orthopedic surgeon. The annotated images were divided into train (1004), validation (174), and test (93) sets. We use *pytorch_segmentation_models* for U-Net implementation and *PyTorch* framework for training the model. The test set was used for final evaluation of the model.

Results: The mean absolute error for CSA and AI between human-performed and machine-performed measurements on the test set with 93 images was 1.68° (95% CI 1.406°–1.979°) and 0.03 (95% CI 0.02 – 0.03), respectively.

Conclusions: A deep learning model can precisely and accurately measure CSA and AI in shoulder anteroposterior radiographs. A tool of this nature makes large-scale research projects feasible and holds promise as a clinical application if integrated with a radiology software program.

© 2022 Mayo Foundation for Medical Education and Research. <https://www.mayoclinic.org/copyright/>. Published by Elsevier Inc. on behalf of American Shoulder & Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Several morphological parameters, such as the anterior acromion morphology, the lateral acromial angle, the coracohumeral interval, the glenoid inclination, the acromion index (AI), and the shoulder critical angle (CSA), have been associated by some authors with the development of rotator cuff tears and glenohumeral osteoarthritis.^{2,3,5,7,10,12}

Nyffeler et al¹² hypothesized that a larger lateral extension of the acromion predisposes the supraspinatus tendon to degeneration due to the increased amount of ascending force of the deltoid muscle, with subsequent impingement of the supraspinatus tendon under the acromion. The AI introduced by Nyffeler et al¹² is the ratio between (1) the distance from the glenoid to the lateral edge of the acromion and (2) the distance from the glenoid to the lateral

edge of the greater tuberosity. Ames et al¹ reported higher association between a larger acromial index and increased rate of rotator cuff tendon tear, and it is also related to more disability by the Quick Disabilities of the Arm, Shoulder and Hand Outcome Measure and poorer physical health as measured by the Short Form-12 Physical Component Summary score. Similarly, the CSA—defined as the angle between the face of the glenoid fossa and a line connecting the glenoid face to the most inferolateral point of the acromion measured on anteroposterior (AP) shoulder radiographs—was introduced by Moor et al¹⁰; smaller and larger CSAs were associated with shoulder osteoarthritis and rotator cuff tears, respectively. In their study, the mean CSA was 33° in healthy shoulders, 28° in patients with osteoarthritis, and 38° in patients with rotator cuff tears.¹⁰

The use of deep learning algorithms for musculoskeletal imaging has increased in recent years.^{8,15,16} To our knowledge, no study has investigated the use of a deep learning model for the prediction of AI or CSA in shoulder radiographs. This study aimed to develop a deep learning tool to automate the measurement of AI and CSA on

Institutional review board approval was not required for this study.

*Corresponding author: Hilal Maradit Kremers, MD, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.

E-mail address: maradit@mayo.edu (H. Maradit Kremers).

<https://doi.org/10.1016/j.xrrt.2022.03.002>

2666-6391/© 2022 Mayo Foundation for Medical Education and Research. <https://www.mayoclinic.org/copyright/>. Published by Elsevier Inc. on behalf of American Shoulder & Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

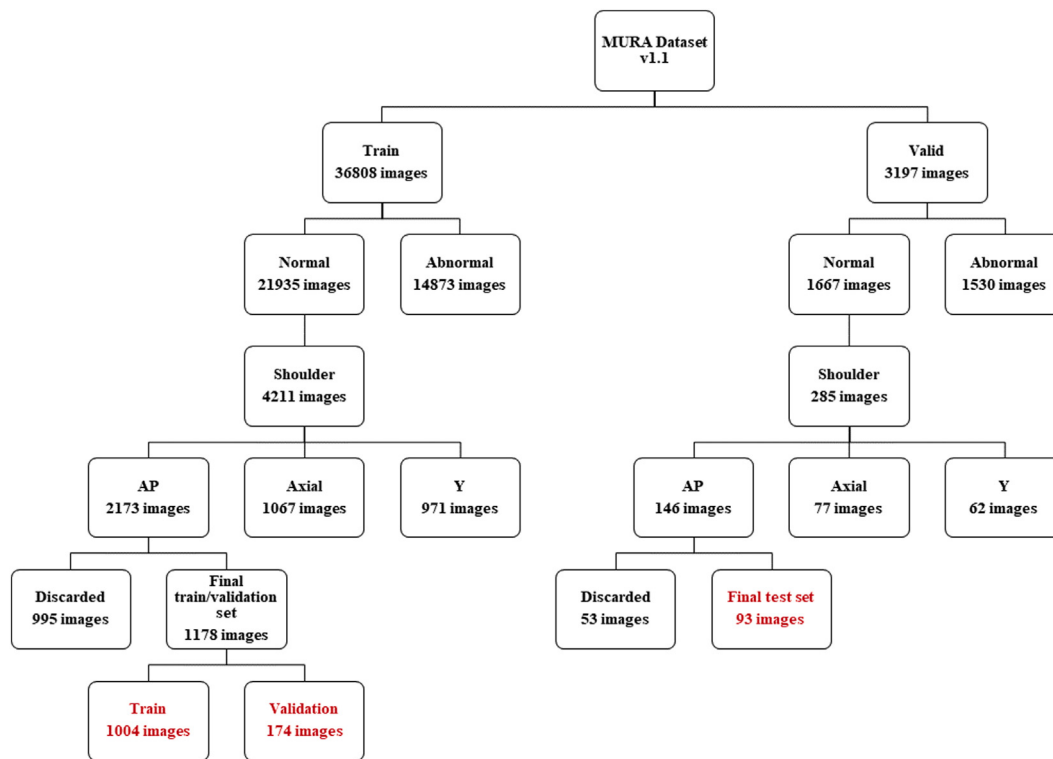


Figure 1 Shoulder AP images collection process. AP, anteroposterior.

AP shoulder radiographs. Our hypothesis was that a convolutional neural network could accurately identify CSA and AI on AP shoulder radiographs. Herein, we introduce a fully automated tool based on a U-Net-like¹⁴ model with EfficientnetB3¹⁸ as its encoder to automatically measure CSA and AI.

Materials and methods

Dataset

In this study, we used the MURA Dataset v1.1, which is a large, publicly available set of musculoskeletal radiographs from the Stanford University School of Medicine (Stanford, CA, USA).¹³ This dataset contains x-ray images of the shoulder, humerus, elbow, forearm, wrist, and hand, labeled as either normal or abnormal. For the purposes of this study, we only used shoulder images labelled as normal. Images were divided into 3 sets for either training, validation, or testing, with no overlap of radiographs between the sets.

Image collection

The MURA Dataset v1.1 categorizes radiographs by patient identification, study number, region, and normal vs. abnormal. Since this dataset is large, we developed a deep learning model to classify the images. To do so, we labeled 195 images (110 training, 85 validation) as AP, axillary, or Y views, which are the most common views used for evaluation of the shoulder joint. Next, we trained a convolutional neural network with 18 layers (ResNet18⁶) on this small subset of images. This simple model fully converged and reached an F1 score of 1.0 for all the 3 views on those 85 validation images. Then, we used this model to quickly classify all the normal shoulder images (4211 images from the dataset's training set and 285 images from the dataset's validation set) into the same 3 classes, which resulted in 2173 AP images, 1067 axillary

images, and 971 Y images from the training set and 146 AP, 77 axillary, and 62 Y images from the validation set.

Image annotation

Next, an experienced orthopedic surgeon annotated 4 landmarks (inferior border of the glenoid, superior border of the glenoid, lateral border of the acromion, and the most lateral part of the proximal humerus) in the normal AP shoulder images using a publicly available image annotation tool (<https://www.robots.ox.ac.uk/~vgg/software/via/via-1.0.6.html>). These landmark annotations were used as ground truth for training the main deep learning model. During this process, we excluded 995 images out of 2173 images in dataset's train subset and 53 out of 146 images in data set's validation subset manually and before starting to train any of the models for the main task of point prediction. These images were excluded due to 1 or more than 1 of the following reasons: (1) Image was not an AP or a true AP image; (2) there was a shoulder prosthesis; (3) the image brightness and quality were impaired.

As such, in the end, we used the remaining 93 images from the dataset's validation set as our final test set, and we divided the remaining 1178 images from the dataset's training set for our own training (1004 images) and validation (174 images). We used our own validation set for hyperparameter optimization, and the results are reported only on the final unseen test set (Fig. 1).

Deep learning model and training strategy

Since predicting the exact coordinates of the target pixels may be suboptimal due to large variances in pixel scale, we use heatmap regression to predict the most probable pixels for each landmark. Therefore, when reading the images, we created 4 corresponding heatmaps for each landmark by replacing the landmark's corresponding pixel in a 2D grid by a Gaussian distribution with standard

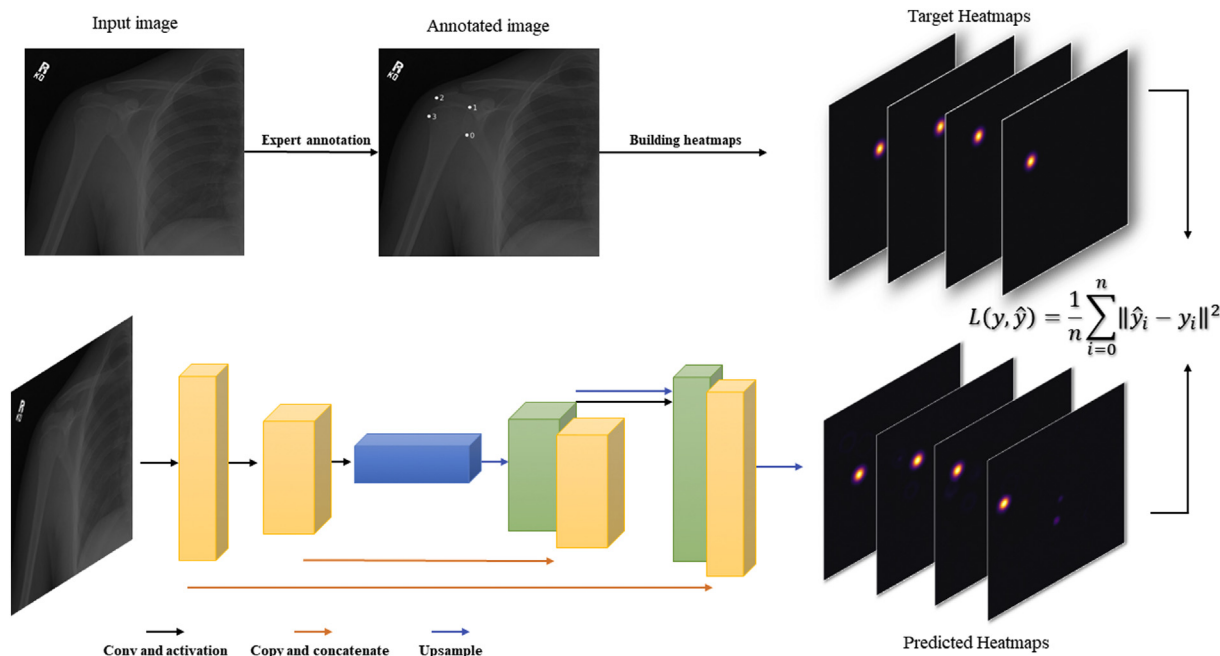


Figure 2 Illustration of data annotation to model training.

deviation of 12. We resized each image to the size 512 * 512, augmented it using data augmentation techniques (see Appendix 1 for more details), and then fed it to a U-Net-like¹⁴ model with EfficientnetB3¹⁸ as its encoder. We used the *pytorch_segmentation_models* library (https://github.com/qubvel/segmentation_models.pytorch) for U-Net implementation and the *PyTorch* framework (<https://pytorch.org>) for training the model. This model predicts a 4-channel output tensor with height and width equal to those of the image (a tensor of shape 4 * 512 * 512). Each channel is the model's prediction for each landmark heatmap. We then computed the loss between model's predictions and the actual heatmaps using the weighted mean squared error (MSE) loss function (see Appendix 1 for more details). Figure 2 provides a summary visualization of data annotation and model training.

After finding the optimal hyperparameters of our model regarding its performance on the validation set (see Appendix 1 for more details), we used cross-validation and trained 5 models on 5 folds of our training and validation set images to best use all the available data. At inference time on the final test set, we first took the average of all the models' predictions to obtain 4 heatmaps per images, and then we obtained the coordinates of each landmark by obtaining the argmax of the corresponding heatmap over the 2 dimensions and then calculated the AI and CSA in a rule-based manner using linear algebra libraries in Python. The 95% confidence intervals (CIs) were calculated with bootstrapping because the distribution of errors was not normal (our code is available at: <https://github.com/moein-shariatnia/MSK>).

Results

For the CSA, the average difference between human-performed and machine-performed measurements on the test set with 93 images was 1.68° (95% CI 1.41°, 1.98°) when determined using the mean absolute error and 2.20° (95% CI 1.85°, 2.55°) when determined using the root MSE. For the AI, the average difference between human-performed and machine-performed measurements on the test set with 93 images was 0.028 (95% CI 0.02, 0.03) when

determined using the mean absolute error and 0.037 (95% CI 0.03, 0.04) when determined using the root MSE (Table 1).

To demonstrate the model's performance and compare it to ground truth labels, Figure 3A provides an illustration of the regions predicted by our model for landmark location. After choosing the pixel with the maximum probability, comparisons with ground truth points can be completed. Then we plot the angle and the lines needed to compute the AI ratio in Figure 3B and C, respectively. Calculating the angle in degree and the ratio is done using rule-based functions by leveraging linear algebra.

Discussion

Osteoarthritis of the glenohumeral joint and rotator cuff tears are common pathologies of the shoulder joint.^{9,11} Although there are several factors that play a role in the etiology of these pathologies, individual morphology of the scapula is one of the factors that have been identified as possibly associated with rotator cuff tears and glenohumeral osteoarthritis. The CSA has been reported by some investigators to correlate with both rotator cuff tendon tears and glenohumeral joint osteoarthritis.

In this study, we developed a fully automated tool to measure the CSA and AI on AP radiographs of the shoulder using a deep learning model. The mean absolute error of our model for CSA measurements was less than 2°. Since the mean CSA reported for normal, rotator cuff tear and glenohumeral osteoarthritis reported by Moor et al¹⁰ had differences of 5° or more, it would be safe to assume that a model with a measurement error less than 2° is a reliable tool with acceptable error level to measure CSA in shoulder

Table 1 CSA and AI predicted by deep learning model.

Target variable	Mean absolute error	Root mean squared error
CSA	1.6848-CI 95% (1.406, 1.979)	2.2018-CI 95% (1.846, 2.548)
AI	0.0275-CI 95% (0.023, 0.033)	0.0368-CI 95% (0.030, 0.043)

CSA, shoulder critical angle; AI, acromion index; CI, confidence interval.

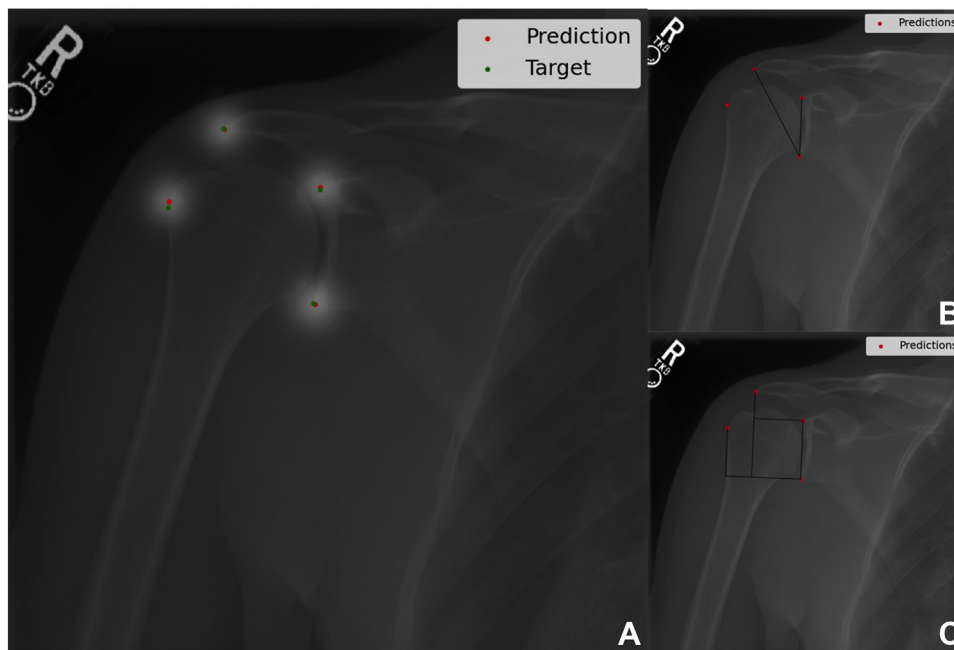


Figure 3 (A) An example of model performance in comparison to the ground truth labels in prediction of landmarks. (B) prediction of CSA by the model. (C) Prediction of AI by the model. CSA, shoulder critical angle; AI, acromion index.

AP images. While the mean absolute error of our model in predicting AI was about 0.03, differences among normal shoulders, shoulders with rotator cuff tears, and glenohumeral osteoarthritis reported by Nyffeler et al¹² were more than 0.1. Therefore, the level of error of our model for AI measurement can be considered acceptable regarding the prediction power of the model in distinguishing different types of pathology in shoulder AP images.

Since all the images were annotated by a single orthopedic surgeon, we cannot report on interobserver reproducibility for human-based CSA and AI measurements. However, we showed an acceptable level of errors in CSA and AI measurements when the model was compared to measurements by an experienced orthopedic surgeon. Previous studies have investigated the reliability of manual CSA and AI measurements. Cherchi et al⁴ reported 96.7% and 95.5% intraobserver and interobserver reproducibility for CSA, respectively. While we could not find exact difference in angles between observers in their manuscript, the interobserver and intraobserver differences for CSA represented in their Bland-Altman plots were about 1° and 2°, respectively. In a study by Spiegl et al,¹⁷ the CSA values measured on radiographs had interobserver and intraobserver agreement of 86.9% and 90.9%, respectively. Moor et al¹⁰ found interclass correlation coefficients of 98% for both CSA and AI. Therefore, our model could be considered an automated algorithm that can accurately and precisely measure these valuable indexes in an AP shoulder image.

There are several potential limitations of this study. First, all images from the MURA dataset were annotated by only 1 orthopedic surgeon and used as ground truth for developing and validation of the deep learning model. Second, the MURA dataset is a deidentified dataset with no demographic information available. This made it impossible to divide the dataset based on age or gender and examine variation of CSA and AI by demographic factors. Third, although we used shoulder images of the MURA dataset that were labeled as normal shoulder AP images, some of the images were not true AP view. Theoretically, the projected CSA could vary according to the position of the scapula, but it did not affect

our measurement model because we did not use it to classify images based on different shoulder pathologies.

Conclusion

We developed a deep learning model that can precisely and accurately measure CSA and AI in shoulder AP images. A tool of this nature makes large-scale research projects feasible and holds promise as a clinical application if integrated with a radiology software program.

Disclaimers:

Funding: This work was supported by the National Institutes of Health (NIH) (grant nos. R01AR73147, P30AR76312, R01AG060920, R01HL147155).

Conflicts of interest: Dr. Sanchez-Sotelo receives royalties and consulting fees from Stryker, consulting fees from Acumed and Exatech, honorarium from JSES, and has stock in PrecisionOS and PSI. Dr. Kremers and the research foundation with which she is affiliated have NIAMS grants (R01AR73147 and P30AR76312). The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.xrrt.2022.03.002>.

References

- Ames JB, Horan MP, Van der Meijden OA, Leake MJ, Millett PJ. Association between acromial index and outcomes following arthroscopic repair of full-thickness rotator cuff tears. *JBS* 2012;94:1862-9. <https://doi.org/10.2106/JBJS.K.01500>.
- Bigliani LU, Ticker JB, Flatow EL, Soslowky LJ, Mow VC. The relationship of acromial architecture to rotator cuff disease. *Clin Sports Med* 1991;10:823-38.

3. Charalambous CP, Eastwood S. Anterior Acromioplasty for the Chronic Impingement Syndrome in the Shoulder: A Preliminary Report. *Classic Papers in Orthopaedics*. Springer; 2014. p. 301-3. https://doi.org/10.1007/978-1-4471-5451-8_74.
4. Cherchi L, Ciernohac J, Godet J, Clavert P, Kempf J-F. Critical shoulder angle: Measurement reproducibility and correlation with rotator cuff tendon tears. *Orthopaedics Traumatol Surg Res* 2016;102:559-62. <https://doi.org/10.1016/j.otsr.2016.03.017>.
5. Davidson PA, Elattrache NS, Jobe CM, Jobe FW. Rotator cuff and posterior-superior glenoid labrum injury associated with increased glenohumeral motion: a new site of impingement. *J Shoulder Elbow Surg* 1995;4:384-90.
6. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015:[arXiv:1512.03385 p.]. Accessed December 10, 2015. Available at: <https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H>.
7. Kim JR, Ryu KJ, Hong IT, Kim BK, Kim JH. Can a high acromion index predict rotator cuff tears? *Int Orthop* 2012;36:1019-24. <https://doi.org/10.1007/s00264-012-1499-4>.
8. Liu F, Kijowski R. Deep learning in musculoskeletal imaging. *Adv Clin Radiol* 2019;1:83-94. <https://doi.org/10.1016/j.yacr.2019.04.013>.
9. Minagawa H, Yamamoto N, Abe H, Fukuda M, Seki N, Kikuchi K, et al. Prevalence of symptomatic and asymptomatic rotator cuff tears in the general population: from mass-screening in one village. *J Orthop* 2013;10:8-12. <https://doi.org/10.1016/j.jor.2013.01.008>.
10. Moor B, Bouaicha S, Rothenfluh D, Sukthankar A, Gerber C. 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. *Bone Joint J* 2013;95:935-41. <https://doi.org/10.1302/0301-620X.95B7.31028>.
11. Nakagawa Y, Hyakuna K, Otani S, Hashitani M, Nakamura T. Epidemiologic study of glenohumeral osteoarthritis with plain radiography. *J Shoulder Elbow Surg* 1999;8:580-4.
12. Nyffeler RW, Werner CM, Sukthankar A, Schmid MR, Gerber C. Association of a large lateral extension of the acromion with rotator cuff tears. *JBJS* 2006;88:800-5. <https://doi.org/10.2106/JBJS.D.03042>.
13. Rajpurkar P, Irvin J, Bagul A, Ding D, Duan T, Mehta H, et al. MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. 2017:[arXiv:1712.06957 p.]. Accessed December 11, 2017. Available at: <https://ui.adsabs.harvard.edu/abs/2017arXiv171206957R>.
14. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015:[arXiv:1505.04597 p.]. Accessed May 18, 2015. Available at: <https://ui.adsabs.harvard.edu/abs/2015arXiv150504597R>.
15. Rouzrokh P, Ramazanian T, Wyles CC, Philbrick KA, Cai JC, Taunton MJ, et al. Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty From Postoperative Radiographs. *J Arthroplasty* 2021. <https://doi.org/10.1016/j.arth.2021.02.028>.
16. Rouzrokh P, Wyles CC, Philbrick KA, Ramazanian T, Weston AD, Cai JC, et al. A Deep Learning Tool for Automated Radiographic Measurement of Acetabular Component Inclination and Version After Total Hip Arthroplasty. *J Arthroplasty* 2021. <https://doi.org/10.1016/j.arth.2021.02.026>.
17. Spiegl UJ, Horan MP, Smith SW, Ho CP, Millett PJ. The critical shoulder angle is associated with rotator cuff tears and shoulder osteoarthritis and is better assessed with radiographs over MRI. *Knee Surg Sports Traumatol Arthrosc* 2016;24:2244-51. <https://doi.org/10.1007/s00167-015-3587-7>.
18. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019:[arXiv:1905.11946 p.]. Accessed May 28, 2019. Available at: <https://ui.adsabs.harvard.edu/abs/2019arXiv190511946T>.