Original Article

Deep Learning for Detecting Supraspinatus Calcific Tendinopathy on Ultrasound Images

CME Credits

Pei-Hsin Chiu[1], Mathieu Boudier-Revéret[2], Shu-Wei Chang[1], Chueh-Hung Wu[3,4]*, Wen-Shiang Chen[3], Levent Özçakar[5]

[1]Department of Civil Engineering, National Taiwan University, Taipei, Taiwan, [2]Department of Physical Medicine and Rehabilitation, Centre Hospitalier de l'Université de Montréal, Montreal, Canada, [3]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan, [4]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan, [5]Department of Physical and Rehabilitation Medicine, Hacettepe University Medical School, Ankara, Turkey

Abstract

Background: The aim of the study was to evaluate the feasibility of convolutional neural network (CNN)-based deep learning (DL) algorithms to dichotomize shoulder ultrasound (US) images with or without supraspinatus calcific tendinopathy (SSCT). **Methods:** This was a retrospective study pertaining to US examinations that had been performed by 18 physiatrists with 3–20 years of experience. 133,619 US images from 7836 consecutive patients who had undergone shoulder US examinations between January 2017 and June 2019 were collected. Only images with longitudinal or transverse views of supraspinatus tendons (SSTs) were included. During the labeling process, two physiatrists with 6-and 10-year experience in musculoskeletal US independently classified the images as with or without SSCT. DenseNet-121, a pre-trained model in CNN, was used to develop a computer-aided system to identify US images of SSTs with and without calcifications. Testing accuracy, sensitivity, and specificity calculated from the confusion matrix was used to evaluate the models. **Results:** A total of 2462 images were used for developing the DL algorithm. The longitudinal-transverse model developed with a CNN-based DL algorithm was better for the diagnosis of SSCT when compared with the longitudinal and transverse models (accuracy: 91.32%, sensitivity: 87.89%, and specificity: 94.74%). **Conclusion:** The developed DL model as a computer-aided system can assist physicians in diagnosing SSCT during the US examination.

Keywords: Artificial intelligence, calcification, machine learning, rotator cuff, sonography

Introduction

Shoulder pain is a common disabling complaint in general medical practice whereby rotator cuff calcific tendinopathy is an upfront diagnosis during the ultrasound (US) examination.[1-3] Since B-mode US is regarded as an excellent imaging tool to visualize calcifications within the rotator cuff tendons,[4,5] it is effectively used to help physicians rationalize the treatment in patients with shoulder pain.[6] However, US is an operator-dependent imaging modality and requires much training/practice to reach a sufficient level of diagnostic accuracy.[7,8] While interobserver agreement is good among experienced physicians, poor agreement is documented when US findings are interpreted by less experienced physicians.[9,10]

To this end, for supporting the decision-making process, computer-aided diagnosis has been developed in recent years. Machine learning (ML) is a part of the broad field of artificial intelligence (AI) and involves systems that aim to construct algorithms which can learn from and make predictions on data.[11] Deep learning (DL), a subclass of methods in the broader field of ML, is a particularly powerful tool for extracting nonlinear features from data. Among DL approaches, convolutional neural networks (CNNs) are the most commonly used ones in the field of medical image analysis.[12] CNN – particularly promising in US – is designed to extract highly representative image features in a fully automatic manner where predictable acoustic features are typically neither obvious nor easily handcrafted.[13]

The unique challenges associated with the application of DL in US are the shortage of images and the expertise required for

Received: 26-09-2021 **Revised:** 08-11-2021 **Accepted:** 15-11-2021 **Available Online:** 16-08-2022

Access this article online

Quick Response Code:

Website: www.jmuonline.org

DOI: 10.4103/jmu.jmu_182_21

Address for correspondence: Dr. Chueh-Hung Wu, No. 2, Sec. 1, Shengyi Road, Zhubei City, Hsinchu County 302, Taiwan.
E-mail: b88401062@ntu.edu.tw, nojred@gmail.com

How to cite this article: Chiu PH, Boudier-Revéret M, Chang SW, Wu CH, Chen WS, Özçakar L. Deep learning for detecting supraspinatus calcific tendinopathy on ultrasound images. J Med Ultrasound 2022;30:196-202.

their acquisition.[14] US has the inherent limitations of inter- and intra-observer variability, which result in discrepancies between the operators in terms of image acquisition and interpretation.[15,16] A sufficient number of US images of adequate quality is required for establishing a DL algorithm having high accuracy. In most previously published articles using DL, the number of US images was <300.[13,17]

Our center is one of the largest musculoskeletal (MSK) US centers in Taiwan whereby services to a large population are provided and a strict long-term US image storage process is maintained. All sonographers in our center are trained with high standards in the same system, and we have an internal review process for image quality and diagnostic accuracy. Accordingly, a large number of normal and abnormal high-quality MSK US images have been accumulated, i.e., adequate for DL development. Given all the aforementioned issues, in this study, we aimed to evaluate the feasibility of CNN-based DL algorithms to dichotomize shoulder US images with or without supraspinatus calcific tendinopathy (SSCT).

Materials and Methods

Data acquisition

Institutional review board approval was obtained for this retrospective study (approval number: 201910110RIND), and the requirement for informed consent for the review of patient images and medical records was waived. All the examinations had been performed by a team of 18 physiatrists (with 3–20 years of experience) who had successfully passed the board examination following an MSK US training program. Some of these examiners have been editors and authors of several MSK US textbooks and some others have been entitled as "Registered in Musculoskeletal® sonography" or "Certified Interventional Pain Sonologist" certifications by the World Institute of Pain. To ensure the heterogeneity of the images for generalizability of this model, US images in the dataset had been obtained from different examiners using various machines including Noblus (Hitachi, Japan), Acuson S2000 (Siemens, Germany), Xario Model SSA-660A (Toshiba, Japan), and Aplio 500 (Canon, Japan).

133,619 US images from 7836 consecutive patients who had undergone shoulder US examinations in our MSK US Center between January 2017 and December 2019 were collected. In our country, the National Health Insurance covers examination fees, and therefore, the referral criteria were relatively liberal. The enrolled patients had been referred for US examinations due to neck and shoulder pain and a standard protocol had been applied to all of them.[18] Only images with longitudinal or transverse views of the supraspinatus tendons (SST) were included for further labeling, whereas those with Doppler US were excluded. In clinical practice, Doppler US image was recorded when there was a high suspicion of abnormality. We excluded these images to prevent the model from "learning" this feature to recognize abnormal images. We aimed to train this model solely based on B-mode images. The number of images of SST in the longitudinal view was 7165 (2222 with calcification and 4943 without) and the number of images in the transverse view was 5201 (1178 with calcification and 4023 without).

During the labeling process, two physiatrists with 6- and 10-year experience in MSK US independently classified the images as with or without SSCT. An image was classified as "with SSCT" when hyper echoic lesions (regardless of their shape, size, or number) with or without acoustic shadowing within the SST were present.[4] The concomitant presence or absence of other findings, such as subdeltoid bursitis, SST tear, or tendinitis, did not contribute to the annotation. If inconsistencies existed between the two physiatrists, a consensus was reached through discussion. If a consensus could not be reached, the image was excluded to ensure quality of the training data.

Data pre-processing

A total of 12366 US images underwent deidentification to secure patients' personal information. Before submitting the images for model training, each image was cropped to remove patient identifications, machine settings, annotations, and scales. Only original US images without any annotations were included in the dataset; US images with annotations that represent the position of calcification labeled by the original sonographers were excluded. The number of cases with and without calcification in the longitudinal view of the SST without labels was 780 and 3784, respectively. Those in the transverse view of the SST without labels were 451 and 2822, respectively. Figure 1 shows the flow chart of data acquisition

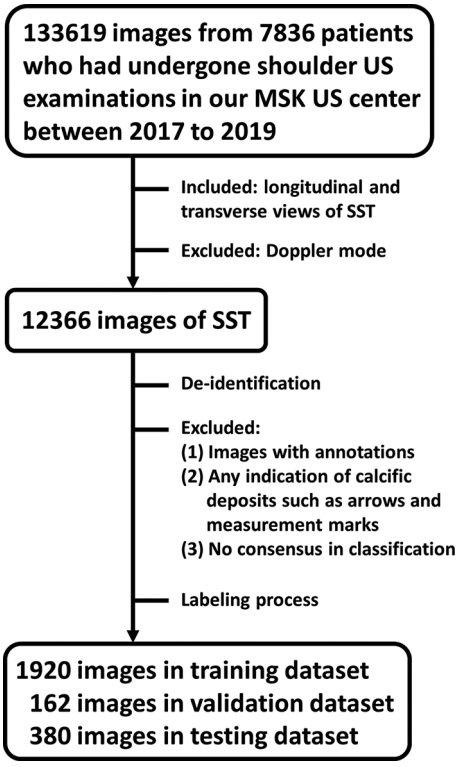

Figure 1: Flow chart of data acquisition and pre-processing. (US: ultrasound; MSK: musculoskeletal; SST: supraspinatus tendon)

and pre-processing. Table 1 gives the training, validation, and testing datasets as regards the US images. Three models, i.e., the longitudinal, transverse, and longi-trans models, were trained and tested with these different datasets. The longitudinal model was trained with only the longitudinal views of the SST images, the transverse model with only the transverse views of the SST images, and the longitrans model with all the SST images.

Data augmentation

Data augmentation was used to diversify US images to decrease the impact of their shortage. All data in the training process were randomly rotated from 0° to 40° to imitate the variation of the position between the SST of each patient. The width shift, height shift, shear, and zoom were randomly set from 0 to 0.20 to imitate the variation originating from the personal variability between physicians. In addition, a channel shift of 10 was used to imitate the color variation between different brands of US machines, and a horizontal flip was used to imitate the variation between different medial/lateral orientations of images.

Model architecture and training

Among all the pretrained networks in the CNN, the models considered in this study were trained with DenseNet-121, which was trained using Image Net, i.e., a dataset containing more than 14 million images over 20,000 classes. The structure of DenseNet-121 was shown in Figure 2. To fit the case in our study, the original top fully connected layer of DenseNet-121 was replaced by two classes for the transverse model, the longitudinal model, and four classes for the longi-trans model as its dense layers. A softmax nonlinearity was followed to show the predicted probability of each class. To avoid overfitting in the models, a dropout layer was used while developing the models. Half of the hidden neurons in the network were randomly deleted when the dropout rate was set to 0.50, and the amount of the input and output neurons were simultaneously kept in each batch. Adam was used as an optimizer with a learning rate of 10^{-5}. Its low memory usage, good computing performance, and the ability to handle noise samples and sparse gradient improve the training process in the models. In addition, categorical cross-entropy was used as the loss function to determine the residual between the ground truth and prediction. To eliminate the effect of imbalanced data between different classes in the longitrans model, the class weight was automatically adjusted. Three final models with the lowest validation loss were selected in different datasets after 300 epochs. Common pretrained networks in classifying medical imaging such as ResNet50 and VGG19 were used to evidence the performance of our model developed with DenseNet-121.

Evaluation index

We used the testing accuracy, sensitivity, and specificity calculated from the confusion matrix to evaluate the models. SST with and without calcification is represented by the positive and negative signs in the confusion matrix, respectively. Of note, the gold standard of diagnosing SSTs with and without calcification involved a consensus between the two experts.

The receiver operating characteristic (ROC) curve was also used as an evaluation index, which is a graphical plot highly

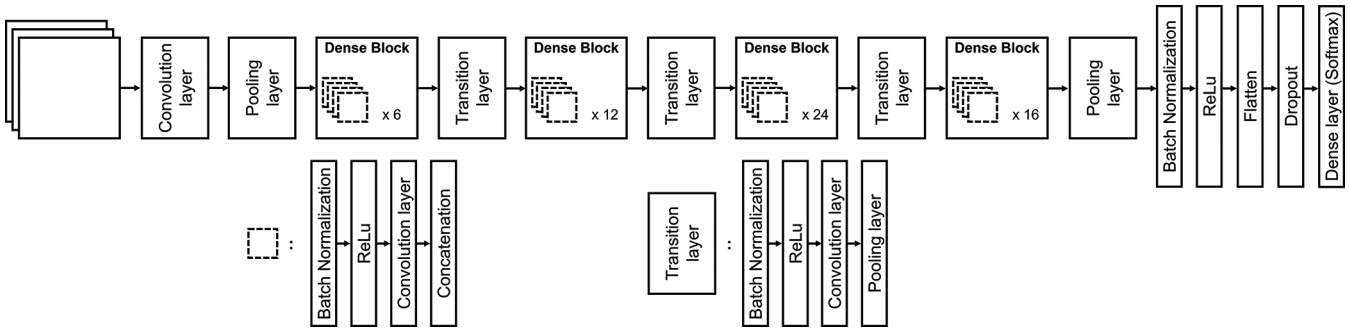

Figure 2: The segmentation frame diagram for the pretrained model, DenseNet-121

Table 1: Distribution of training, validation, and testing datasets in the different models (*n*)

	Longitudinal model		Transverse model		Longi-trans model		Testing dataset		
	Training dataset	Validation dataset	Training dataset	Validation dataset	Training dataset	Validation dataset	Testing dataset (a)	Testing dataset (b)	Testing dataset (c)
All	1200	100	720	62	1920	162	260	120	380
Cal (+), L	600	50	0	0	600	50	130	0	130
Cal (−), L	600	50	0	0	600	50	130	0	130
Cal (+), T	0	0	360	31	360	31	0	60	60
Cal (−), T	0	0	360	31	360	31	0	60	60

All: Total number of ultrasound images, Cal (+): Calcification present, Cal (−): Calcification absent, L: Longitudinal view of supraspinatus tendons, T: Transverse view of supraspinatus tendons

correlated with the confusion matrix. The false-positive rate and true positive rate with the threshold change were plotted as ROC curves. The more the ROC curve skewed to the upper left side of the coordinate chart, the better the model. When the area under the ROC curve area under the curve (AUC) is > 0.90, 0.80–0.90, and 0.70–0.80, the model has outstanding, excellent, and acceptable discrimination, respectively. However, models with AUC <0.50 have no discrimination.[19]

The decision-making process of ML is called into question due to the "black boxes" which usually represent data that go in and results that go out without the master of the process for human. To open the black boxes and obtain approval from a human being, heatmaps were used in ML to visualize what the model "saw" in the input images. Heatmaps are types of data visualization tools that depict data using different colors in two dimensions. In this study, Gradient-weighted Class Activation Mapping[20] was used to plot heatmaps. It calculates weights through back propagation and multiplies weights with feature maps to gain the importance of regions in the last convolutional layer of the CNN.

Results

Figure 3a and b show the accuracy and loss curves of the longitudinal model, respectively. The training and validation accuracies of the longitudinal model were 99.50% and 96.88%, respectively. Its training and validation losses were 0.01 and 0.13 at epoch 257, respectively. For the transverse model, the training and validation accuracies were 99.86% and 94.64%, respectively. Its training and validation losses at epoch 339 were 0.01 and 0.20, respectively. Figure 3c and d show the accuracy and loss curves of the transverse model. For the longi-trans model, the training and validation accuracies were 96.72% and 90.62%, respectively. The training and validation losses at epoch 313 were 0.03 and 0.41, respectively. Its accuracy and loss curves are shown in Figure 3e and f.

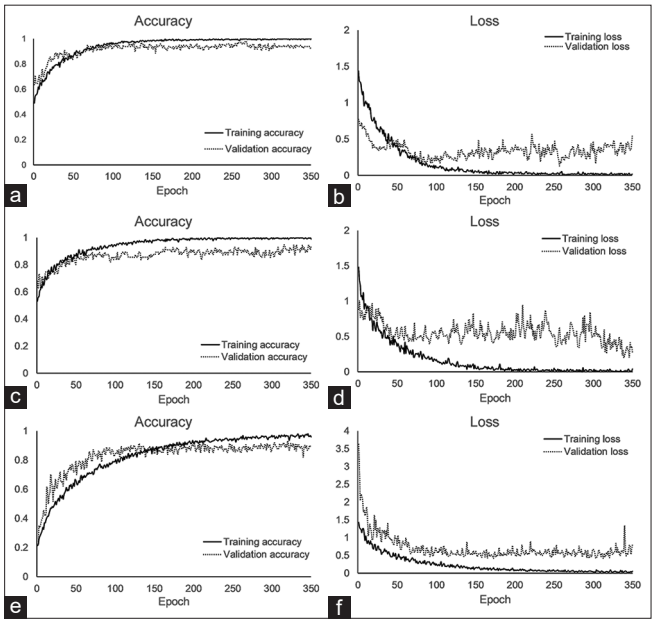

Figure 3: Learning curve of the different models with their accuracy and loss. (a) Accuracy curve of the longitudinal model; (b) Loss curve of the longitudinal model; (c) Accuracy curve of the transverse model; (d) Loss curve of the transverse model; (e) Accuracy curve of the longi-trans model; (f) Loss curve of the longi-trans model

Figure 4a is the testing accuracy of the three models developed with DenseNet-121 against different testing datasets. The testing accuracy of the longitudinal model testing against the longitudinal view of SST was 92.31%. The testing accuracy of the transverse model testing against the transverse view of SST was 89.17%. Herewith, these models had lower testing accuracies when they were tested against an untrained view of SST, i.e., 78.33% for the longitudinal model tested against the transverse view of SST and 69.23% for the transverse model tested against the longitudinal view of SST. Further, the longitrans model performed better based on its testing accuracies, which were 91.54% for the longitudinal view of SST and 90.83% for the transverse view of SST. As such, models trained on the longitudinal and transverse views of SST had a more accurate diagnosis, although models trained on single views of SST also had the ability to a certain extent. Moreover, multiple views of SST improved the quantity/richness of input data rather than confusing the diagnostic criteria of the models.

Figure 4b shows the sensitivity of the three models developed with DenseNet-121 against different testing datasets. No obvious difference was found in their sensitivities, which were approximately 83% to 89%. However, their specificity did not show the same findings, i.e., specificity against trained views of SST was approximately 95% among the three models [Figure 4c]. On the contrary, the specificity against untrained views of SST was 25% to 45% lower than that of the trained views of SST.

Related comparison in models developed with ResNet50 and VGG19 are shown in Figure 4d-i. Even though some of the index such as the testing accuracy of transverse model developed with ResNet50 against the longi-trans testing dataset was higher than that of model developed with DenseNet-121, its corresponding sensitivity did not show in the same way. Moreover, models developed with DenseNet-121 presented the highest composition of testing accuracy, sensitivity, and specificity between the three ML pretrained models. Therefore, models developed with DenseNet-121 were analysed in the further step.

The AUCs of the longitudinal model testing against testing datasets (a), (b), and (c) were 0.95, 0.89, and 0.93, respectively. The AUCs of the transverse model tested against different testing datasets were 0.80, 0.93, and 0.84. The AUCs of the longi-trans model against the three testing datasets were 0.95, 0.94, and 0.95.

Figure 5 shows the heatmaps for different views of the SST with or without calcification. All the heatmaps indicated that the occurrence or non-occurrence of calcification in the

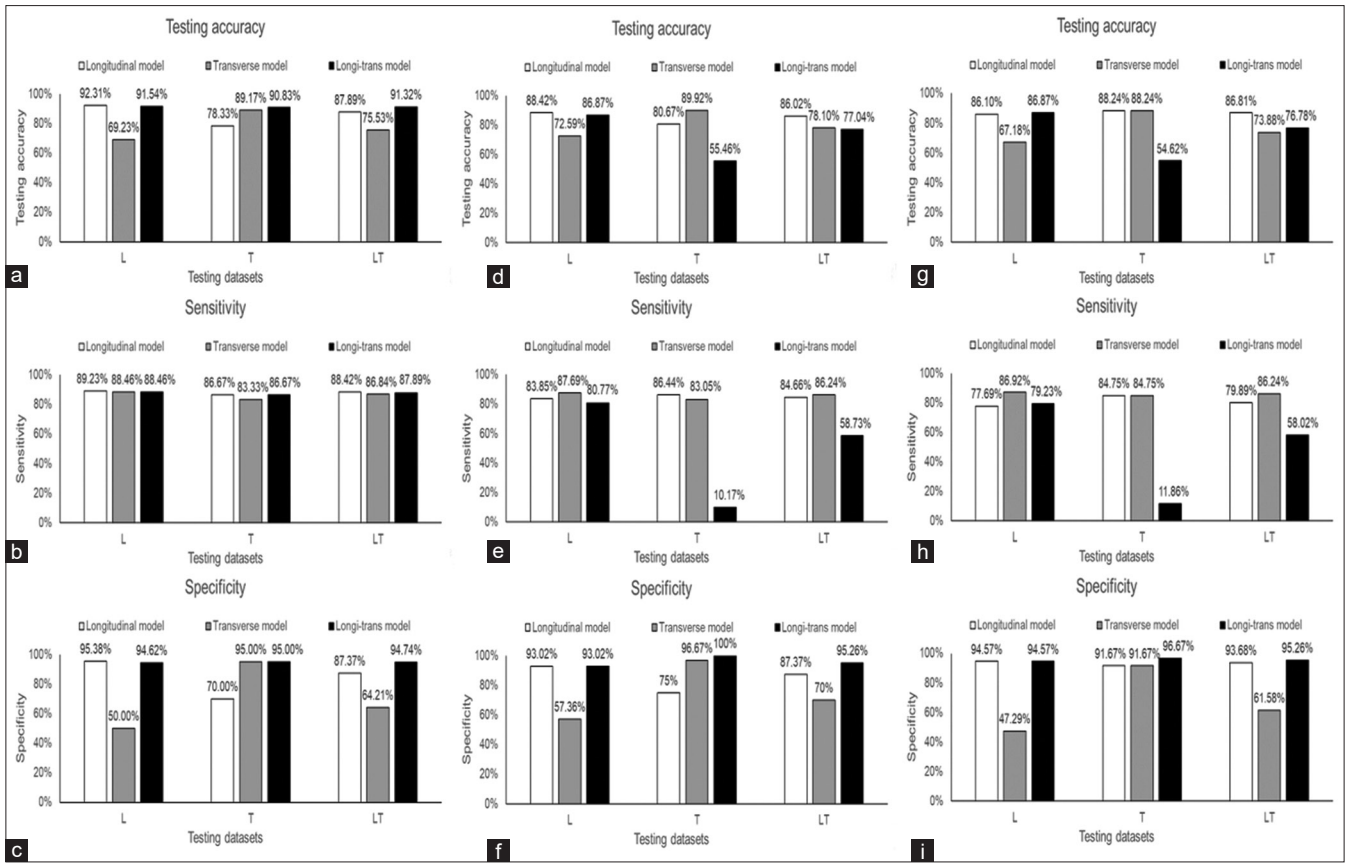

Figure 4: Evaluation indices of three models developed with three pre-trained models against different testing datasets (L: longitudinal view of supraspinatus tendons, $N = 260$; T: transverse view of supraspinatus tendons, $N = 120$; LT: combination of the above two datasets, $N = 380$). (a-c)Testing accuracy, sensitivity, specificity of models developed with DenseNet121. (d-f) Testing accuracy, sensitivity, specificity of models developed with ResNet50. (g-i) Testing accuracy, sensitivity, specificity of models developed with VGG19

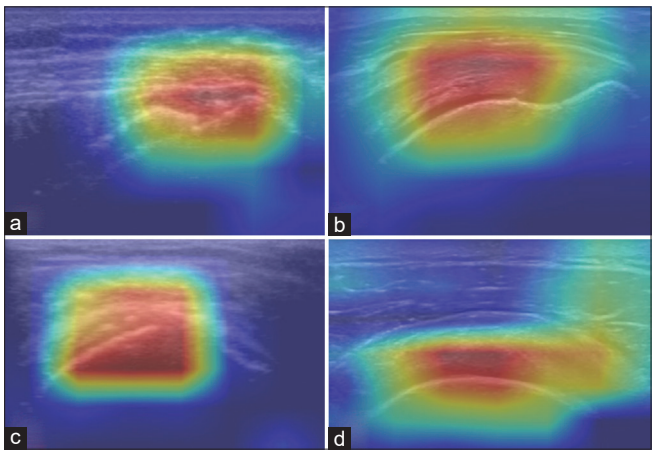

Figure 5: Heatmaps for the supraspinatus tendon. Longitudinal view of the supraspinatus tendon (a) with calcification and (b) without calcification. Transverse view of the supraspinatus tendon (c) with calcification and (d) without calcification

SST was a key parameter for the algorithm to determine the existence of SSCT. Moreover, whether the view of the SST was longitudinal or transverse did not affect the algorithm results. Figure 6 shows several examples of US images that were correctly and incorrectly classified by the algorithm.

Discussion

In this study, we tried to develop a DL algorithm that can detect SSCT on US images. Among the three models trained on different views of SST, the longi-trans model trained with the longitudinal and transverse views of SST had a better performance in diagnosing the existence of SSCT. Although the accuracy and loss of the longitrans model during the training and validation stages were worse than those of the longitudinal and transverse models, its performance against unseen testing data was more credible and much better than that of the other two models, especially when simultaneously diagnosing the longitudinal and transverse views of SSCT. In practice, it is not convenient and often not possible to use only one view of SST to diagnose SSCT. Therefore, the ability to diagnose the longitudinal and transverse views of SSCT is indispensable. The longi-trans model tested on more than 300 images was proven to correctly diagnose the existence of calcification, not only in the longitudinal but also in the transverse view of SST.

Comparing the performance of our longitrans model to the other four applications of ML in US images, the longi-trans model seems to have a much better performance. In the pertinent literature, a model used for the segmentation and

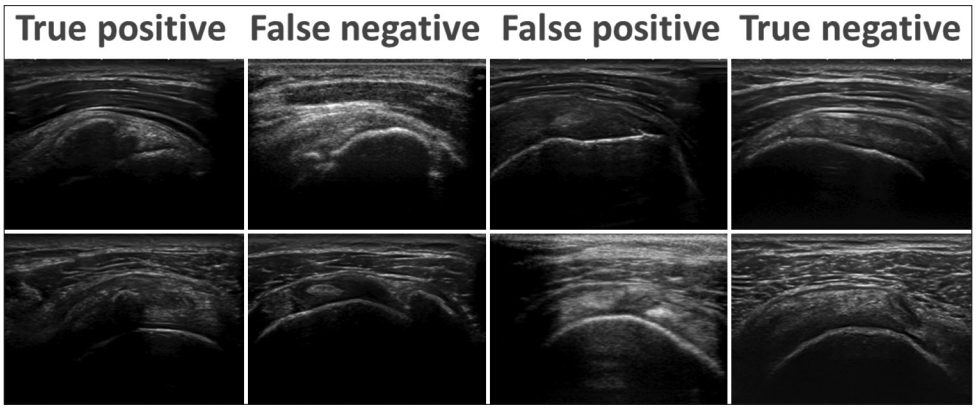

Figure 6: Correctly and incorrectly classified examples of ultrasound images. Longitudinal (upper row) and transverse (lower row) views of the supraspinatus tendon

classification of breast lesions was initially developed with 80.42% accuracy.[21] Second, a model classifying diffuse liver diseases was developed with 82.6% accuracy.[22] Thereafter, Guo and Du[23] developed a model for the classification of thyroid US standard planes and it had an accuracy of 83.88%. In addition, the model for the classification of breast cancer US images was developed with 88% accuracy.[24] In short, our longi-trans model outperforms other ML applications concerning US images.

It is noteworthy that US is an operator-dependent imaging modality, and variable interoperator agreements have been observed even among experienced sonographers.[9,10] For those who have little experience in diagnosing SSCT, a computer-aided diagnostic tool may help/improve the diagnosis. In the near future, these algorithms may analyze – either in real time or very shortly after data extraction – the US images generated by clinicians or technologists and assist them in performing accurate diagnoses. Needless to say, this would make US an even-more attractive first-line imaging modality, as long as the images generated are standardized to a certain level.

As another clinical implication of this algorithm, SSCT can be applied to augment medical education. In medical students, Cheng *et al*.[25] reported that AI-assisted learning helps in achieving a significantly higher diagnostic accuracy of hip fractures on pelvic radiographs as compared to those who undergo conventional training. Although it was not the primary aim of our study, we believe that an increase in the diagnostic accuracy of this CNN-based DL algorithm can enable its application to the design of training software that, for example, can be incorporated in commercialized US machines.

Our study has several limitations. First, it was a retrospective study conducted in a single institution and this might limit the generalizability of its results. All examiners followed a standard protocol for scanning, but different physiatrists used different US machines and machine settings. Moreover, while rising the amount of US images through the data augmentation step, the fixed parameters may not rise the generalizability of increased data used in model building. We believe that the results, to a certain extent, revealed relevant 'real-world' heterogeneity, which was highly important for training a DL algorithm.[26] One future research direction would be to verify the diagnostic accuracy of this algorithm using a test dataset obtained also from other US centers. Second, it is important to note that we did not consider the morphologic characteristics of the calcific deposits e. g. size, shape, and echogenicity. One inherent limitation of applying DL in US is the shortage of US images. Although US images were increased through data augmentation step before model building, the data increased by the fixed parameters could hardly replace the real ones. To perform subgroup analysis, the number of images for each subgroup of calcific deposits should be sufficient to divide into training and validation set with representative amount which avoids models from overfitting. Pooling together US images with all types of calcific deposits would yield a significant amount of heterogeneity. Third, no radiographic correlation was performed in this study. Nonetheless, US is a well-established tool for the diagnosis of SSCTs[4,27,28] and, in our hospital, patients with shoulder pain are not routinely referred to take plain films, partly due to the fear of radiation. Moreover, since this study involved retrospective image analyses, it would be unethical (and quite impossible) to perform shoulder radiography in the study population. Fourth, this model only classified US images as "with calcification" and "without calcification." There was no information regarding whether other abnormalities-such as SST tears, tendinosis or tendinitis and sub acromial bursitis-existed in those images. A future model may be developed to perform multiclass classifications. For example, the model can classify one image as SST tear with SSCT, and another as sub acromial bursitis with SSCT.

Conclusion

The longi-trans model-developed with a CNN-based DL algorithm-was better for the diagnosis of SSCT than the longitudinal and transverse models, with an accuracy of 91.32%, sensitivity of 87.89%, and specificity of 94.74%. Our

results seem to be promising for the clinical application of DL algorithms to establish diagnoses based on MSK US images.

Acknowledgments

We appreciate Ms. Wan-Ju Chien for her assistance in collection of the medical images.

Financial support and sponsorship

The research was funded by National Taiwan University Hospital (109-M4670) and Ministry of Science and Technology, Taiwan (MOST 107-2314-B-002-045-MY3).

Conflicts of interest

Prof. Wen-Shiang Chen, an editorial board member at *Journal of Medical Ultrasound*, had no roles in the peer review process of or decision to publish this article. The other authors declared no conflicts of interest in writing this paper.

References

1. Feleus A, Bierma-Zeinstra SM, Miedema HS, Bernsen RM, Verhaar JA, Koes BW. Incidence of non-traumatic complaints of arm, neck and shoulder in general practice. Man Ther 2008;13:426-33.
2. Ottenheijm RP, van't Klooster IG, Starmans LM, Vanderdood K, de Bie RA, Dinant GJ, *et al.* Ultrasound-diagnosed disorders in shoulder patients in daily general practice: A retrospective observational study. BMC Fam Pract 2014;15:115.
3. Louwerens JK, Sierevelt IN, van Hove RP, van den Bekerom MP, van Noort A. Prevalence of calcific deposits within the rotator cuff tendons in adults with and without subacromial pain syndrome: Clinical and radiologic analysis of 1219 patients. J Shoulder Elbow Surg 2015;24:1588-93.
4. Farin PU, Jaroma H. Sonographic findings of rotator cuff calcifications. J Ultrasound Med 1995;14:7-14.
5. Papatheodorou A, Ellinas P, Takis F, Tsanis A, Maris I, Batakis N. US of the shoulder: Rotator cuff and non-rotator cuff disorders. Radiographics 2006;26:e23.
6. Ottenheijm RP, Cals JW, Weijers R, Vanderdood K, de Bie RA, Dinant GJ. Ultrasound imaging for tailored treatment of patients with acute shoulder pain. Ann Fam Med 2015;13:53-5.
7. Iossifidis A, Ibrahim EF, Petrou C. Ultrasound for the detection of full-thickness rotator cuff tears: The learning curve for an orthopaedic surgeon using a novel training method. Shoulder Elbow 2015;7:158-62.
8. McCormack RA, Nayyar S, Jazrawi L. Physician training: Ultrasound and accuracy of diagnosis in rotator cuff tears. Bull NYU Hosp Jt Dis 2016;74:207.
9. O'Connor PJ, Rankine J, Gibbon WW, Richardson A, Winter F, Miller JH. Interobserver variation in sonography of the painful shoulder. J Clin Ultrasound 2005;33:53-6.
10. Ohrndorf S, Naumann L, Grundey J, Scheel T, Scheel AK, Werner C, *et al.* Is musculoskeletal ultrasonography an operator-dependent method or a fast and reliably teachable diagnostic tool? Interreader agreements of three ultrasonographers with different training levels. Int J Rheumatol 2010;2010:164518.
11. Erickson BJ. Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging Radiographics 2017;37:505-15.
12. Fedewa R, Puri R, Fleischman E, Lee J, Prabhu D, Wilson DL, *et al.* Artificial Intelligence in Intracoronary Imaging. Curr Cardiol Rep 2020;22:46.
13. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: Status, methods, and future opportunities. Abdom Radiol (NY) 2018;43:786-99.
14. Huang Q, Zhang F, Li X. Machine learning in ultrasound computer-aided diagnostic systems: A survey. Biomed Res Int 2018;2018:5137904.
15. Ye FY, Lyu GR, Li SQ, You JH, Wang KJ, Cai ML, *et al.* Diagnostic performance of ultrasound computer-aided diagnosis software compared with that of radiologists with different levels of expertise for thyroid malignancy: A multicenter prospective study. Ultrasound Med Biol 2021;47:114-24.
16. Chang RF, Lee CC, Lo CM. Computer-aided diagnosis of different rotator cuff lesions using shoulder musculoskeletal ultrasound. Ultrasound Med Biol 2016;42:2315-22.
17. Shin Y, Yang J, Lee YH, Kim S. Artificial intelligence in musculoskeletal ultrasound imaging. Ultrasonography 2021;40:30-44.
18. Ozcakar L, Kara M, Chang KV, Tekin L, Hung CY, Ulaülı AM, *et al.* EURO-MUSCULUS/USPRM basic scanning protocols for shoulder. Eur J Phys Rehabil Med 2015;51:491-6.
19. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 2000.
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision.Venice, Italy; IEEE. 2017.
21. Liang Y, He R, Li Y, Wang Z. Simultaneous segmentation and classification of breast lesions from ultrasound images using Mask R-CNN. In: 2019 IEEE International Ultrasonics Symposium (IUS). Glasgow, UK: IEEE; 2019.
22. Dandan L, Huanhuan M, Xiang L, Yu J, Jing J, Yi S. Classification of diffuse liver diseases based on ultrasound images with multimodal features. In: 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). Auckland, New Zealand: IEEE; 2019.
23. Guo M, Du Y. Classification of thyroid ultrasound standard plane images using resnet-18 networks. In: 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID). Xiamen, China: IEEE; 2019.
24. Latif G, Butt MO, Yousif Al Anezi F, Alghazo J. Ultrasound image despeckling and detection of breast cancer using deep CNN. In: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF). Ho Chi Minh City, Vietnam: IEEE; 2020.
25. Cheng CT, Chen CC, Fu CY, Chaou CH, Wu YT, Hsu CP, *et al.* Artificial intelligence-based education assists medical students' interpretation of hip fracture. Insights Imaging 2020;11:119.
26. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, *et al.* Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020;368:l6927.
27. Le Goff B, Berthelot JM, Guillot P, Glémarec J, Maugars Y. Assessment of calcific tendonitis of rotator cuff by ultrasonography: Comparison between symptomatic and asymptomatic shoulders. Joint Bone Spine 2010;77:258-63.
28. Chianca V, Albano D, Messina C, Midiri F, Mauri G, Aliprandi A, *et al.* Rotator cuff calcific tendinopathy: From diagnosis to treatment. Acta Biomed 2018;89:186-96.