Medical & Biological Engineering & Computing (2022) 60:1269–1278
https://doi.org/10.1007/s11517-022-02502-6

ORIGINAL ARTICLE

Classification of rotator cuff tears in ultrasound images using deep learning models

Thao Thi Ho[1] · Geun-Tae Kim[2] · Taewoo Kim[1] · Sanghun Choi[1] · Eun-Kee Park[3]

Received: 8 September 2021 / Accepted: 31 December 2021 / Published online: 18 January 2022

Abstract

Rotator cuff tears (RCTs) are one of the most common shoulder injuries, which are typically diagnosed using relatively expensive and time-consuming diagnostic imaging tests such as magnetic resonance imaging or computed tomography. Deep learning algorithms are increasingly used to analyze medical images, but they have not been used to identify RCTs with ultrasound images. The aim of this study is to develop an approach to automatically classify RCTs and provide visualization of tear location using ultrasound images and convolutional neural networks (CNNs). The proposed method was developed using transfer learning and fine-tuning with five pre-trained deep models (VGG19, InceptionV3, Xception, ResNet50, and DenseNet121). The Bayesian optimization method was also used to optimize hyperparameters of the CNN models. A total of 194 ultrasound images from Kosin University Gospel Hospital were used to train and test the CNN models by five-fold cross-validation. Among the five models, DenseNet121 demonstrated the best classification performance with 88.2% accuracy, 93.8% sensitivity, 83.6% specificity, and AUC score of 0.832. A gradient-weighted class activation mapping (Grad-CAM) highlighted the sensitive features in the learning process on ultrasound images. The proposed approach demonstrates the feasibility of using deep learning and ultrasound images to assist RCTs' diagnosis.

Keywords Rotator cuff tears · Deep learning · Convolutional neural network · Ultrasound · Transfer learning

Abbreviations

AUC	Area under the curve
CNN	Convolution neural network
CT	Computed tomography
DL	Deep learning
FN	False negative
FP	False positive
Grad-CAM	Gradient-weighted class activation mapping
MRI	Magnetic resonance imaging
RCTs	Rotator cuff tears
ROC	Receiver operating characteristic
TN	True negative
TP	True positive

Thao Thi Ho and Geun-Tae Kim contributed equally.

✉ Sanghun Choi
s-choi@knu.ac.kr

Eun-Kee Park
eunkee.park@gmail.com

[1] School of Mechanical Engineering, College of Engineering, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea

[2] Department of Internal Medicine, College of Medicine, Kosin University, 262 Gamcheon-ro, Seo-gu, Busan 49267, Republic of Korea

[3] Department of Medical Humanities and Social Medicine, College of Medicine, Kosin University, 262 Gamcheon-ro, Seo-gu, Busan 49267, Republic of Korea

1 Introduction

Rotator cuff tears (RCTs) are the major cause of musculoskeletal pain in the shoulders [1], and millions of people are suffering from RCT injuries in the world [1, 2]. Trauma and degeneration are the two main causes of rotator cuff injuries [3]. RCTs are most likely to occur in the supraspinatus muscle and tendon [3, 4]. These injuries mostly occur among middle-aged and older individuals, being associated with a reduction of their quality of life. This is possibly due to a decreased range of motion in the shoulder joint with pain. Over time, partial-thickness tears enlarge and spread into full-thickness tears and develop distinct chronic pathological changes due to muscle retraction, fatty infiltration, and muscle atrophy [5, 6]. The challenge lies in the early diagnosis of RCTs because the majority of patients have no or marginal symptoms. Rotator cuff problems are often subjectively assessed by qualitative or semi-quantitative methods [3, 7, 8]. Despite being fast and easy to use, the effectiveness of these methods has been reported to be temporary and unreliable [9, 10] due to the complexity and subjectivity of classification systems. An accurate diagnosis of RCTs is primarily based on a combination of patient's symptoms and medical techniques [11].

For more robust and accurate diagnosis, medical imaging methods are recently used. First, quantitative computed tomography (CT) and magnetic resonance imaging (MRI) can provide excellent visualization of the structural details with high spatial resolution but have not yet been widely adopted in increasingly busy clinical workflows, mainly because of time constraints and high cost [8, 12–14]. Next, MRI has a higher contrast resolution for the evaluation of soft tissues, and CT enables detailed quantitative analysis of the shoulder using specific density values [8, 12] compared with ultrasound images. In the meantime, ultrasound is often used to diagnose musculoskeletal diseases with advantages over CT or MRI such as safe, low cost, portable, and tolerance by patients [15–18]. In addition, ultrasound is commonly used as a stethoscope of musculoskeletal diseases in most rheumatology clinics, because it can objectify physical examination by directly comparing physical examination findings with ultrasound findings and it can be displayed directly to patients and explained in real-time [16]. Currently, shoulder ultrasound images are increasingly being used in the imaging evaluation of RCTs [18, 19]. Once the ultrasound data of a patient's shoulder is obtained, a skilled doctor easily locates the tears from several longitudinal or transverse images and determines the presence, absence, and size of the tears as well as the necessity of a surgical operation.

In recent years, deep learning (DL) has been utilized for great success in the field of computer vision including classification, object detection, and segmentation. The approach also has been used for a diagnosis of muscle state in CT and MRI datasets [20–27], detection of large RCTs from conventional shoulder radiographs [28], and segmentation RCTs for diseased regions in ultrasound images [29]. With the significant increase in hardware performance and availability of big data, DL-based image processing methods have outperformed traditional methods. Therefore, an automated DL-based diagnosis is desirable to be used for enhancing clinicians' diagnostic performance in terms of accuracy and speed.

To our best knowledge, DL techniques have not been studied much for diagnosing RCTs, which are not clearly and easily identifiable in ultrasound images. As a result, this study aims at developing and evaluating the performance of a convolution neural network (CNN) for automatically classifying RCTs from ultrasound shoulder images. Our proposed approach uses 2D information of an ultrasound shoulder slice (longitudinal or transverse views) and performs the entire RCTs diagnosis immediately without a segmentation step. Only final diagnostic information (RCTs or non-RCTs) and resize images are used for pre-processing. Especially, the layers and hyperparameters of pre-trained models are tuned using Bayesian optimization [30]. Furthermore, we also use gradient-based class activation map (Grad-CAM) [31] method to visualize the feature extraction in the learning process of CNN. Consequently, the proposed method automatically determines the RCTs size in two categories (normal and tear) and visualizes the localization information of the RCTs as a second opinion for clinical decisions.

2 Methods

This study was approved by the institutional review boards of Kosin University Gospel Hospital (KUGH 2020–07-020) and the requirement for patient consent was waived.

2.1 Study population

In this study, we initially collected data from 194 ultrasound images of 103 subjects (87 subjects with both longitudinal and transverse views, 6 subjects with longitudinal view, and 10 subjects with transverse view), which were categorized into RCT and non-RCT groups. All shoulder ultrasound scans have been collected at Kosin University Gospel Hospital by ALOKA Ultrasound Alpha 7 machine. The detailed statistics about the demographics of subjects with RCTs and non-RCTs are summarized in Table 1. For the RCTs data group, we used 72 subjects (142 ultrasound images). The

Table 1 Demographic and clinical data for event vs. event-free dataset

	Total ($N=103$)	RCTs group ($N=72$)	Non-RCTs group ($N=31$)	P value
Age (years), mean (SD)	59.6 (12.9)	0.037	55.3 (14.8)	<0.273
Sex, n (%)				<0.489
Female	67 (65.0)	48 (66.7)	19 (61.3)	
Male	36 (35.0)	24 (33.3)	12 (38.7)	

mean (*SD*) age of this group was 61.1 (11.8) years with 24 males and 48 females. For the non-RCTs data group, we had 31 subjects (52 ultrasound images) with 12 males and 19 females and a mean age of 55.3 (14.8) years.

2.2 Pre-processing

All ultrasound images were saved in a PNG format. We cropped and resized the images to 224 × 224 pixels to match the original size of the pre-trained models (Fig. 1). Next, the pixel value was normalized to a range between 0 and 1 using the MinMax scaler to prevent the models from overfitting for both training and testing datasets.

2.3 D deep learning models

The detailed structure of the classification network is shown in Fig. 2 for the classification of RCTs based on abdomen ultrasound images. This network was developed using transfer learning based on one of the following 5 pre-trained CNN models: VGG19 [32], ResNet50 [33], InceptionV3 [34], DenseNet121 [35], and Xception [36]. Transfer learning is widely used in the field of DL, because it allows to utilize the advantage of deep networks and develop an accurate model in a short time with comparatively little data [37]. In transfer learning, the learned knowledge of an already trained model is applied to a different but related problem [37]. Based on our data size and to reduce the training time, we used predefined weights (on ImageNet) for each pre-trained CNN model and updated them through learning process for our dataset to classify RCTs.

After the pre-processing, the grayscale imaging data (1 channel) was copied to the 3 RGB channels (224 × 224 × 3 pixels) in the input layer to fit with the original size of pre-trained models. This is a necessary step to feeding grayscale data into the pre-trained model layers. In the pre-trained model layers, we included each of the five pre-trained models (VGG19, ResNet50, InceptionV3, DenseNet121, and Xception; trained on ImageNet dataset). Each pre-trained network is usually composed of two parts: convolutional base and classifier [38]. The convolutional base includes many convolutional and pooling layers to perform feature extraction from images. The classifier classifies input images based on the extracted features by the convolutional base. Here, we only kept the convolutional base in its original form as a feature extractor and replaced the classifier with a new classifier for RCTs and non-RCTs.

2.4 Automatic tuning method

Traditional fine-tuning is usually used for CNN models to get the desired results. However, they are tuned manually and there is no definite procedure to calculate these hyperparameter values. In this study, the optimal number of CNN layers and hyperparameters were tuned using Bayesian optimization to improve model performance [30]. Bayesian optimization is a sequential model-based optimization algorithm that aims to learn the structure of the hyperparameter space of a black box that is costly to evaluate [30]. By using hyperparameters that appear from previous results to decide the next hyperparameter value, Bayesian optimization is more efficient in time and memory capacity for tuning many hyperparameters than both grid search and random search [39]. There are two components in Bayesian optimization methods. The first component is a probabilistic model of the loss function, i.e., a function that takes the values of the hyperparameters as input and estimates the value of the loss the corresponding

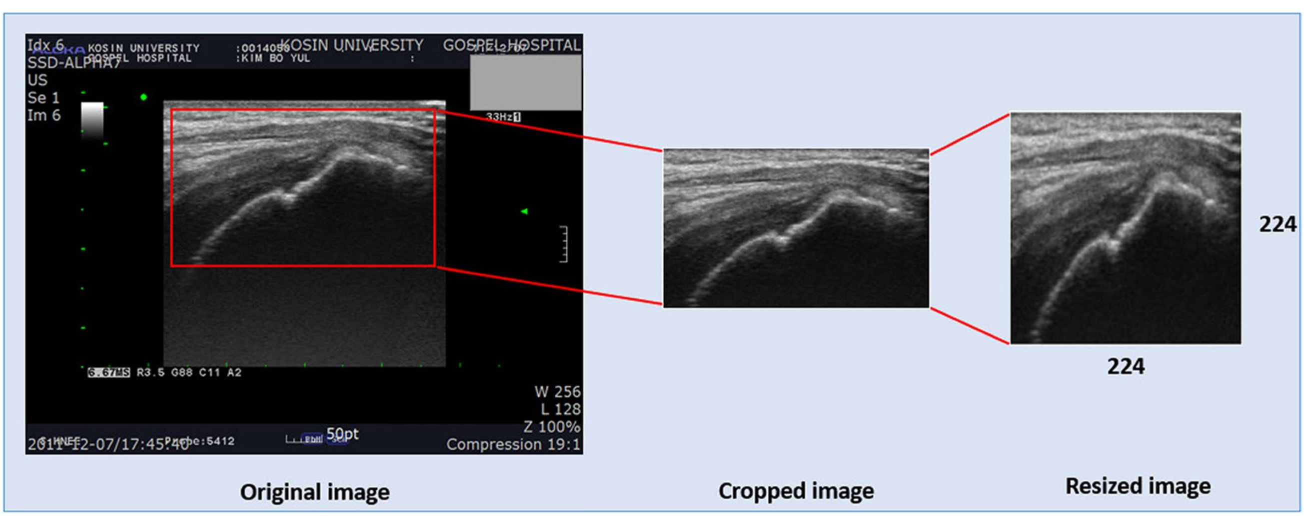

Fig. 1 The representative ultrasound image and the detailed pre-processing step

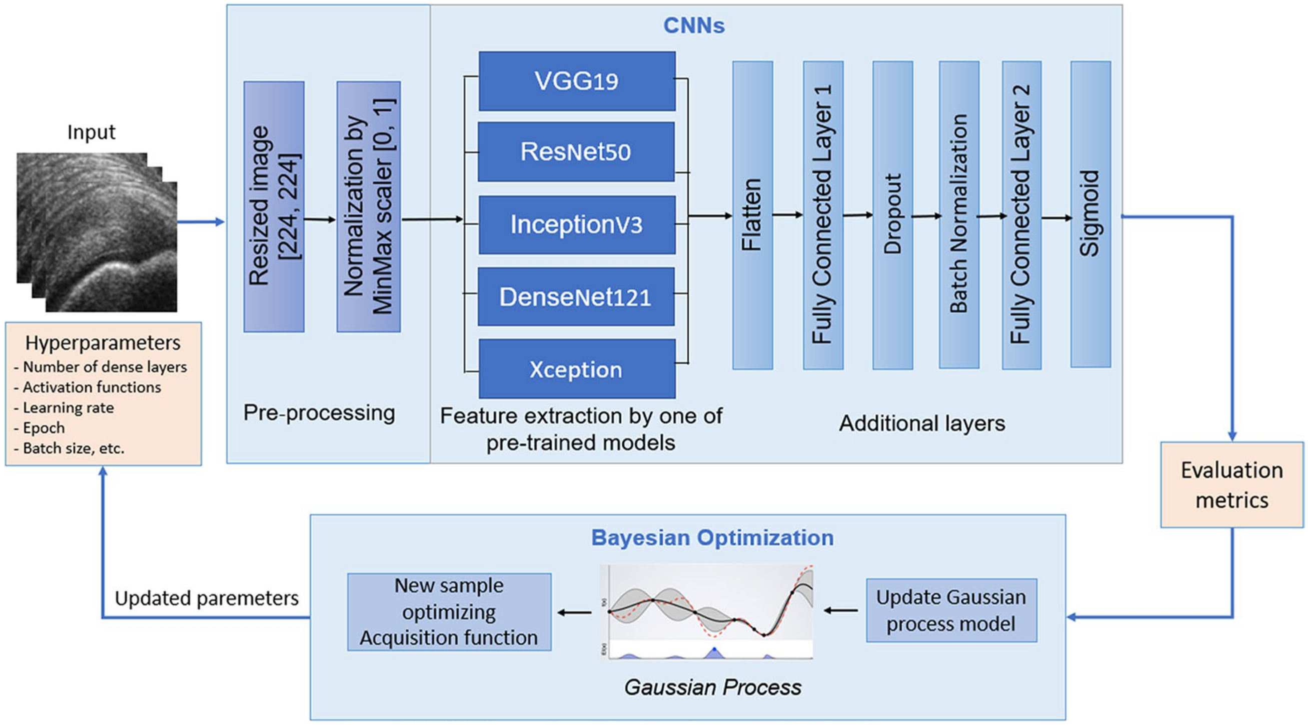

Fig. 2 The procedure of our study with pre-trained models with Bayesian optimization

Table 2 Description of the hyperparameters with their range and optimal value

Hyperparameters	Range	Best results
Number of dense layers	[1; 4]	2
Number of neurons for dense layer	[64: 512]	256
Dropout rate	[0; 0.6]	0.2
Activation in each layer	[Relu, sigmoid, softplus, tanh, selu, elu, exponential]	Relu
Optimizer	[Adam, SDG, RMSprop, Adagrad, Adamax, Nadam, Adadelta]	Adam
Learning rate	[0.0001; 0.1]	0.001
Epoch	[20, 2500]	500
Batch size	[8; 64]	50

neural network would have. Gaussian processes are the typical choice. The second component, called the acquisition function, samples the model of the loss function to select the next set of hyperparameters to evaluate [30]. The hyperparameters to tune are the number of neurons, activation function, optimizer, learning rate, batch size, and epochs. The second step is to tune the number of layers. A detail of the hyperparameter space with their range is shown in Table 2. The range of the number of convolution layers is {1, 2, 3, 4} and the number of neurons is {64, 128, 256, 512}. We implemented dropout to generalize the network performance after dense layer with a rate in range {0.0, 0.6}. There are 7 activation functions and 7 optimizers to tune in this step. The learning rate in the optimizer is also tune with the range {0.0001; 0.1}. The two last tuned hyperparameters are batch size and number of epochs with the range {8, 64} and {20, 2500}, respectively.

2.5 Optimization results

The obtained values from Bayesian optimization are shown in Table 2. In the additional layers of each pre-trained model, we first flattened activations from pre-trained model layers and connected with a fully connected layer. It has 256 neurons for the first dense layer. Batch normalization and dropout layers with a dropout rate of 0.2 were utilized to accelerate training, increase the generalization, and avoid overfitting. The outputs from second fully connected layer were fed into a sigmoid layer, which provided the probability for each RCTs and non-RCTs prediction. During the training, Adam optimizer was used with a learning rate of 0.001. Each pre-trained model was trained with binary cross-entropy loss and 500 epochs. The training batch size is 50 by experiments.

2.6 Validation of AI models and statistical analysis

All image data were trained with five-folds cross-validation into a training set and a test set. The training set was then further separated into sets used for model training and internal validation. The test set was used only for independent testing of the trained models. We trained our pre-trained models using an Intel, Xeon (E5-2640) CPU, and a NVIDIA GeForce RTX 2080Ti. We evaluated the classification performance for five pre-trained models (VGG19, ResNet50, InceptionV3, DenseNet121, and Xception) using sensitivity, specificity, accuracy, precision, and F1-score [40]. These metrics were calculated using true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) based on the RCTs and non-RCTs, as follows:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Sensitivity = TP/(TP + FN) \quad (3)$$

$$Specificity = TN/(TN + FP) \quad (4)$$

$$F1_{score} = 2.\frac{Precision.Sensitivity}{Precision + Sensitivity} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

We then applied Grad-CAM to visualize key learning features from ultrasound data by DenseNet121 model. Grad-CAM created a 2D heatmap as an explanation of the CNN learning process. In addition, we plotted the receiver operating characteristic (ROC) curve and computed the area under the curve (AUC) for each pre-trained model. The statistical comparison of the demographic data was performed using the Mann–Whitney *U*-test for age and the chi-square test for gender.

3 Results

Figure 3 shows the accuracies and losses on the training and validation datasets of five pre-trained models. For all five models, the training and validation losses decreased continuously until the end of training, reaching approximately 0 after 500 epochs; meanwhile, the accuracies increased gradually to more than ~98% for training and ~90% for validation after 500 epochs. Compared with other models, the DenseNet121 model showed the best convergence in both validation loss and accuracy.

The performance of the pre-trained models for classifying RCTs and non-RCTs is summarized in Table 3. Based on the test data, the model of DenseNet121 showed the highest accuracy (88.2%), over InceptionV3, ResNet50, Xception, and VGG19. However, Xception achieved the highest sensitivity (98.6%) and specificity (90.8%) compared to other models. The training time of DL models was relatively short, ranging from 2419 to 4791 s depending on the number of parameters of each model. InceptionV3 model completed the training the earliest with 2419 s, followed by DenseNet121 with 2891 s and Xception with the longest training time of 4791 s.

The ROC curves and confusion matrices for the pre-trained models are presented in Figs. 4 and 5. It is clear from Fig. 5 that 14 RCT subjects were misclassified non-RCT subjects, and 9 non-RCT subjects were misclassified as RCTs by DenseNet121. VGG19 and ResNet50 had more FP cases than InceptionV3. However, just only 6 non-RCT subjects were wrongly predicted as RCTs by ResNet50 compared with 9 subjects by DenseNet121. More specifically, DenseNet121 showed TP and TN of 133/142 and 38/52, respectively. VGG19 showed TP and TN of 134/142 and 32/52, respectively. With ResNet50, these ratios were 136/142 and 29/52, respectively (Fig. 5). For 5 models DenseNet121, InceptionV3, ResNet50, VGG19, and Xception, the values of AUC were 0.832, 0.802, 0.779, 0.845, and 0.753, respectively (Fig. 4).

To evaluate the Grad-CAM visualization results, we validated whether the Grad-CAM location coincides with a shoulder specialist's RCTs marking. Figure 6 shows the DenseNet121-discovered sensitive muscle areas in ultrasound images with brighter colors. It was confirmed that Grad-CAM locations were the same as the actual tear locations for RCT subjects (Fig. 6A). Note that the green arrow in Fig. 6A was identified by an experienced medical doctor.

4 Discussion

In this paper, we described a 2D DL approach based on muscle ultrasound images with two different longitudinal and transverse views as an automatically diagnostic aid capable of detecting RCTs with high accuracy. The proposed method is based on transfer learning from pre-trained models with fine-tuning. The numerical experiment showed that DenseNet121 outperforms the other pre-trained models in binary classification performance: 88.2% accuracy, 90.1% precision, 93.8% sensitivity, 83.6% specificity, and AUC score of 0.832. In addition, the proposed method took less than 1.4 h (with a GPU NVIDIA GeForce RTX 2080Ti) to train models and 3 s to diagnose test data, while still enabling accurate RCTs' diagnosis. The advantage of DenseNet121 created by DenseBlock, which the feature map sequentially reuses throughout the whole network. Another advantage of the proposed method is that pre-processing of the training data is simple; no further pre-processing for ultrasound images such as segmentation was performed. The 2D ground truth ultrasound images were labeled with only minimal information: RCT or non-RCT subjects for training. The proposed models not

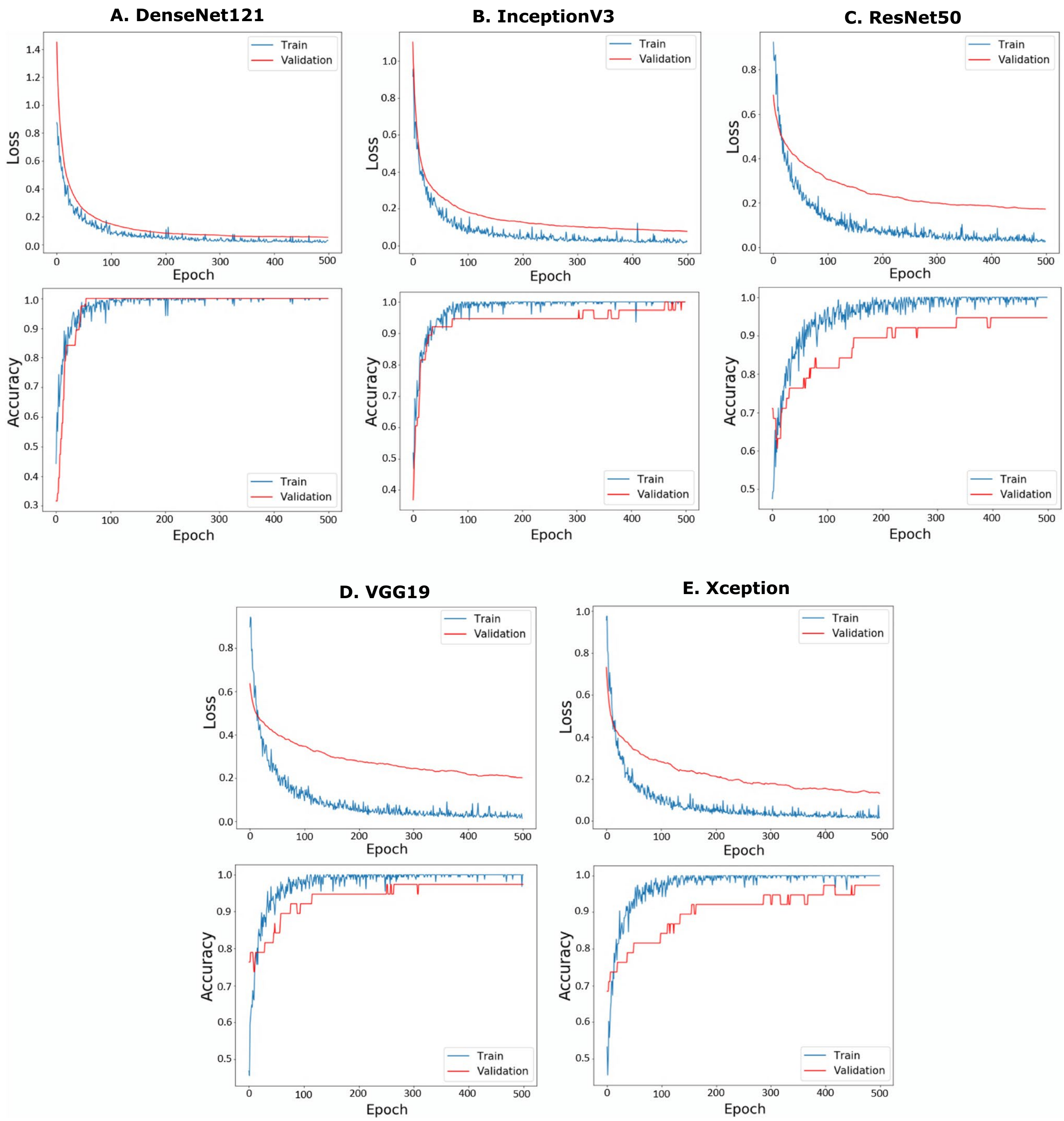

Fig. 3 Plots of loss and accuracy of the training and validation datasets for 5 pre-trained models

Table 3 Performance of pre-trained models and training time for prediction of RCTs versus non-RCTs images

Model	Accuracy	Precision	Sensitivity	F1	Specificity	AUC score	Training time (s)
DenseNet121	88.2	90.1	93.8	92.1	83.6	0.832	2891
InceptionV3	86.1	86.7	97.2	91.3	90.4	0.802	2419
ResNet50	85.1	85.5	95.8	90.4	84.3	0.779	3249
VGG19	85.6	87.7	94.4	90.8	70.5	0.845	3429
Xception	85.1	84.2	98.6	90.8	90.8	0.753	4791

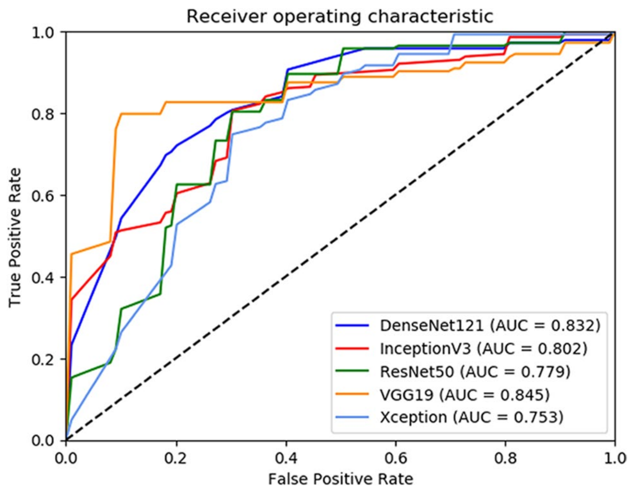

Fig. 4 ROC curves of five pre-trained models

only provide diagnostic results but also the information about location and shape of the tears in each ultrasound image by Grad-CAM method. This supervised DL method is easy to use while providing accurate clinical results, and 2D visualized localization information about a shoulder injury. We confirmed that DL models also could use ultrasound images to classify rotator cuff tear subjects.

Table 4 overviews the datasets, methods, and classification results of some recent DL studies that distinguish RCTs from non-RCTs. From Table 4, different studies use different types of data and DL methods for RCTs classification tasks. Kim et al. [28] proposed a 2D CNN model using shoulder radiographs concatenated with clinical information for rule out significant RCTs in patients suspected of RCTs. They obtained 97.3% sensitivity but the specificity only 45.5%. Kim et al. [25] used MRI scans and a 3D CNN model to

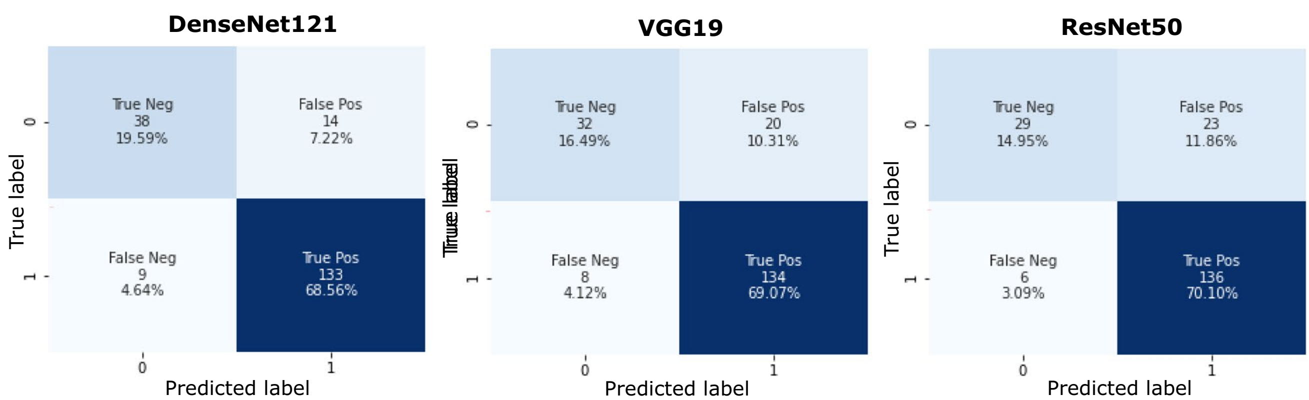

Fig. 5 Confusion matrices of DenseNet121, VGG19, and ResNet50

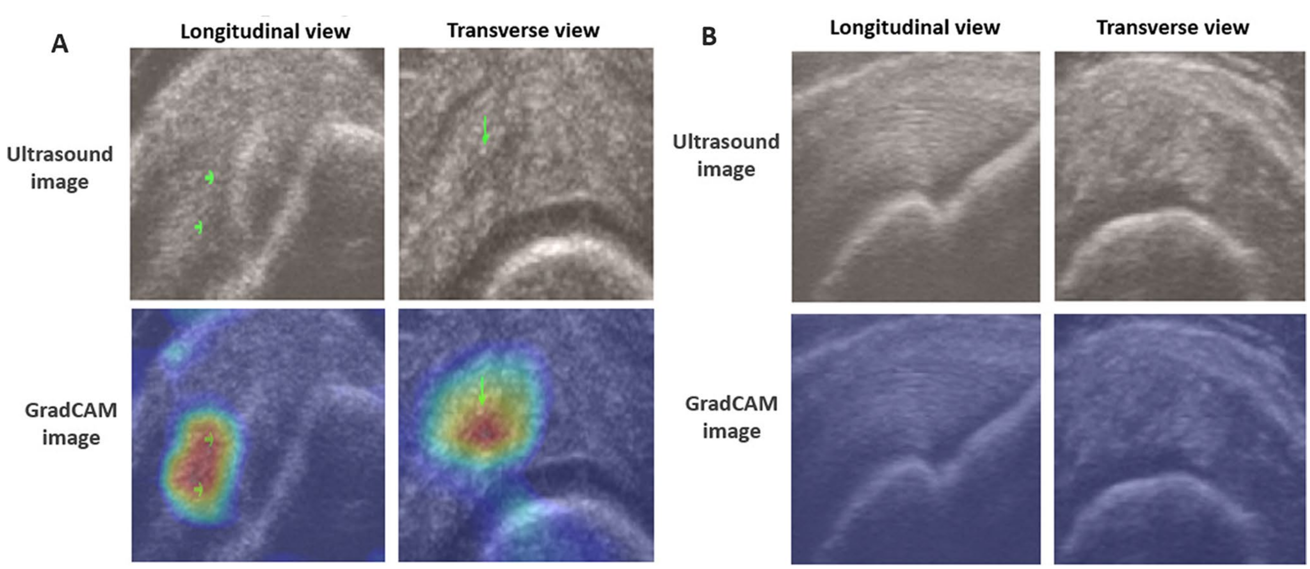

Fig. 6 The representative case for **A** label 1 (RCTs) and **B** label 0 (non-RCTs), and their Grad-CAM heatmap images. Each case included longitudinal and transverse views

Table 4 Previously published classification results of RCTs vs. non-RCTs datasets

Related work	Data	Methods	Performance (%)
Kim et al. [28]	Shoulder radiographs and clinical information ($n=6793$)	2D CNN concatenated with clinical information	Sensitivity 97.3% and specificity 45.5%
Kim et al. [25]	MRI scans ($n=1963$)	3D CNN	Accuracy 87.0% and AUC score of 0.970
Shim et al. [24]	MRI scans ($n=2124$)	3D CNN	Accuracy 92.5%, sensitivity 94.0%, and specificity 90.0%
Lu el al. [26]	Clinical information ($n=169$)	ANN	Accuracy 94.0%, sensitivity 87.0%, and specificity 95.0%
Lee et al. [29]	Ultrasound images ($n=1400$)	CNN segmentation with modified loss function	Accuracy 92.6%, precision 60.4%, and sensitivity 94.2%
Our work	Ultrasound images ($n=194$)	Pre-trained CNN	Accuracy 88.2%, precision 90.1%, sensitivity 93.8%, specificity 83.6%, and AUC score of 0.832

diagnose RCTs. Their method achieved 87.0% accuracy and an AUC score of 0.970. Shim et al. [24] also constructed a novel 3D CNN model with MRI scans to diagnose the tear size with 92.5% accuracy, 94.0% sensitivity, and 90.0% specificity. Lu et al. [26] proposed an ANN model with only clinical data. This ANN model showed a good performance with 94.0% accuracy, 87.0% sensitivity, and 95.0% specificity. Medina et al. [41] introduced a combination CNNs to select a specific sagittal image (Y-view) image and segment RCTs using MRI. Lee et al. [29] developed a novel DL structure to segment a location of RCTs with imbalanced and ultrasound images by modified loss function. Their method achieved 92.6% accuracy, 60.4% precision, and sensitivity of 94.2%. The above methods used labeled segmentation images by human participation during processing or training. This segmentation step is time-consuming and requires the experience and skill of physician who has used ultrasound. Our proposed method obtained 88.2% accuracy, 90.1% precision, 93.8% sensitivity, 83.6% specificity, and AUC score of 0.832, which is comparable to the studies above. The advantages of our proposed approach are its simplicity and computational speed. With these results, our method can assist in making real-time diagnoses with RCTs, especially for some less experienced clinicians or medical facilities that lack modern diagnostic imaging equipment. The automation of ultrasound techniques would make it clinically viable and could further promote the use of ultrasound as the priority imaging technique for rotator cuff injuries.

This study has few limitations. First, our approach gathered data from just one center and one imaging modality with a relatively lower proportion of non-RCT subjects. This reflects the population-based prevalence of RCTs in our data collection area. Considering the same proportion as real-life data may enhance the reliability of the current deep learning models. The test dataset was obtained from the same sources as the training dataset. Thus, this approach needs to be validated in a different dataset obtained from different imaging modalities. Second, we used a relatively small amount of imaging data to train the DL models, even though statistical power analysis was performed in this study. We also used data augmentation to increase amount of data by rotation angle 15° from 0 to 180° and flipped along *x*- and *y*-axes, the accuracy results did not increase too much. Thus, we decided to train CNN models without data augmentation to save time. In conclusion, transfer learning with pre-trained CNN models can be used to construct effective classifiers for ultrasound images. We plan to improve our models by using more ultrasound images and performing more complex tasks (segmentation RCTs and classifying three different grades of RCTs or the tear size or distinguishing RCTs from other osteoarthritis-related diseases) to provide quantitative information in future works.

5 Conclusion

In summary, this study developed and evaluated the performance of pre-trained CNNs for automatically classifying RCTs from ultrasound shoulder images. Our proposed approach used 2D information of an ultrasound shoulder slice (longitudinal or transverse views) and performed the entire RCTs diagnosis immediately without a segmentation step. Only final diagnostic information (RCTs or non-RCTs) and resize images were used for pre-processing. Especially, the layers and hyperparameters of pre-trained models were tuned using Bayesian optimization. Consequently, the proposed method could automatically determine the RCTs size in two categories (normal and tear) and visualize the localization information of the RCTs as a second opinion for clinical decisions. Grad-CAM locations were found to be the same as the actual tear locations identified by a shoulder specialist's RCTs. The DL results in this study indicated that DL has the potential to classify and diagnose muscle-related diseases from relatively low spatial resolution ultrasound images.

Acknowledgements This work was supported by the Korea Ministry of Environment (MOE) as "The Environmental Health Action Program" [2018001360004], and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [NRF-2018R1D1A1B07040886, and NRF-2021R1F1A1060436].

Author contribution TTH, TK, SC, G-TK, and E-KP designed the experiments and interpreted the results. G-TK, E-KP, and SC collected experimental data. TTH, SC, and TK performed the experiments. TTH, E-KP, and SC performed the analyses and wrote the manuscript. E-KP, and SC served as co-corresponding authors. All the authors provided feedback on the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Via AG, De Cupis M, Spoliti M, Oliva F (2013) Clinical and biological aspects of rotator cuff tears. Muscles Ligaments Tendons J. 3(2):70–9. https://doi.org/10.11138/mltj/2013.3.2.070.
2. Craig R, Holt T, Rees JL (2017) Acute rotator cuff tears. BMJ 11(359):j5366. https://doi.org/10.1136/bmj.j5366
3. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC (1994) Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. Clin Orthop Relat Res 304:78–83. https://doi.org/10.1097/00003086-199407000-00014
4. Lapner PL, Jiang L, Zhang T, Athwal GS (2015) Rotator cuff fatty infiltration and atrophy are associated with functional outcomes in anatomic shoulder arthroplasty. Clin Orthop Relat Res 473(2):674–682. https://doi.org/10.1007/s11999-014-3963-5
5. Gladstone JN, Bishop JY, Lo IK, Flatow EL (2007) Fatty infiltration and atrophy of the rotator cuff do not improve after rotator cuff repair and correlate with poor functional outcome. Am J Sports Med 35(5):719–728. https://doi.org/10.1177/0363546506297539
6. Donohue KW, Ricchetti ET, Ho JC, Iannotti JP (2018) The association between rotator cuff muscle fatty infiltration and glenoid morphology in glenohumeral osteoarthritis. J Bone Joint Surg Am. 100(5):381–7. https://doi.org/10.2106/JBJS.17.00232
7. Thomazeau H, Rolland Y, Lucas C, Duval JM, Langlais F (1996) Atrophy of the supraspinatus belly. Assessment by MRI in 55 patients with rotator cuff pathology. Acta Orthop Scand 67(3):264–8. https://doi.org/10.3109/17453679608994685
8. van de Sande MA, Stoel BC, Obermann WR, Tjong a Lieng JG, Rozing PM. (2005) Quantitative assessment of fatty degeneration in rotator cuff muscles determined with computed tomography. Invest Radiol 40(5):313–319. https://doi.org/10.1097/01.rli.0000160014.16577.86
9. Oh JH, Kim SH, Choi JA, Kim Y, Oh CH (2010) Reliability of the grading system for fatty degeneration of rotator cuff muscles. Clin Orthop Relat Res 468(6):1558–1564. https://doi.org/10.1007/s11999-009-0818-6
10. Slabaugh MA, Friel NA, Karas V, Romeo AA, Verma NN, Cole BJ (2012) Interobserver and intraobserver reliability of the Goutallier classification using magnetic resonance imaging: proposal of a simplified classification system to increase reliability. Am J Sports Med 40(8):1728–1734. https://doi.org/10.1177/0363546512452714
11. McElvany MD, McGoldrick E, Gee AO, Neradilek MB, Matsen FA 3rd (2015) Rotator cuff repair: published evidence on factors associated with repair integrity and clinical outcome. Am J Sports Med 43(2):491–500. https://doi.org/10.1177/0363546514529644
12. Terrier A, Ston J, Dewarrat A, Becce F, Farron A (2017) A semi-automated quantitative CT method for measuring rotator cuff muscle degeneration in shoulders with primary osteoarthritis. Orthop Traumatol Surg Res 103(2):151–157. https://doi.org/10.1016/j.otsr.2016.12.006
13. Park J, Kim S, Lim JK, Jin KN, Yang MS, Chae KJ et al (1985) (2021) Quantitative CT image-based structural and functional changes during asthma acute exacerbations. J Appl Physiol 131(3):1056–1066. https://doi.org/10.1152/japplphysiol.00743.2020
14. Yoon S, Tam TM, Rajaraman PK, Lin CL, Tawhai M, Hoffman EA et al (1985) (2020) An integrated 1D breathing lung simulation with relative hysteresis of airway structure and regional pressure for healthy and asthmatic human lungs. J Appl Physiol 129(4):732–747. https://doi.org/10.1152/japplphysiol.00176.2020
15. Klibanov AL, Hossack JA (2015) Ultrasound in radiology: from anatomic, functional, molecular imaging to drug delivery and image-guided therapy. Invest Radiol 50(9):657–670. https://doi.org/10.1097/RLI.0000000000000188
16. Grassi W, Filippucci E, Busilacchi P (2004) Musculoskeletal ultrasound. Best Pract Res Clin Rheumatol 18(6):813–826. https://doi.org/10.1016/j.berh.2004.05.001
17. Tan A (2003) Imaging of the musculoskeletal system: magnetic resonance imaging, ultrasonography and computed tomography. Best Pract Res Clin Rheumatol 17(3):513–528. https://doi.org/10.1016/s1521-6942(03)00021-4
18. Saraya S, El Bakry R (2016) Ultrasound: can it replace MRI in the evaluation of the rotator cuff tears? Egypt J Radiol Nucl Med 47(1):193–201. https://doi.org/10.1016/j.ejrnm.2015.11.010
19. de Jesus JO, Parker L, Frangos AJ, Nazarian LN (2009) Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol 192(6):1701–1707. https://doi.org/10.2214/AJR.08.1241
20. Hashimoto F, Kakimoto A, Ota N, Ito S, Nishizawa S (2019) Automated segmentation of 2D low-dose CT images of the psoas-major muscle using deep convolutional neural networks. Radiol Phys Technol 12(2):210–215. https://doi.org/10.1007/s12194-019-00512-y
21. Weber KA, Smith AC, Wasielewski M, Eghtesad K, Upadhyayula PA, Wintermark M et al (2019) Deep learning convolutional neural networks for the automatic quantification of muscle fat infiltration following whiplash injury. Sci Rep 9(1):7973. https://doi.org/10.1038/s41598-019-44416-8
22. Burns JE, Yao J, Chalhoub D, Chen JJ, Summers RM (2020) A machine learning algorithm to estimate sarcopenia on abdominal CT. Acad Radiol 27(3):311–320. https://doi.org/10.1016/j.acra.2019.03.011
23. Graffy PM, Liu J, Pickhardt PJ, Burns JE, Yao J, Summers RM (2019) Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment. Br J Radiol 92(1100):20190327. https://doi.org/10.1259/bjr.20190327
24. Shim E, Kim JY, Yoon JP, Ki SY, Lho T, Kim Y et al (2020) Automated rotator cuff tear classification using 3D convolutional neural network. Sci Rep 10(1):15632. https://doi.org/10.1038/s41598-020-72357-0
25. Kim M, Park Hm, Kim JY, Kim SH, Hoeke S, Neve WD (2020) MRI-based diagnosis of rotator cuff tears using deep learning and weighted linear combinations. In: Proceedings of the 5th Machine Learning for Healthcare Conference. 292–308
26. Lu HY, Huang CY, Su CT, Lin CC (2014) Predicting rotator cuff tears using data mining and Bayesian likelihood ratios. PLoS ONE 9(4):e94917. https://doi.org/10.1371/journal.pone.0094917
27. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML (2017) Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group no. 132. Med Phys 44(7):e43–e76. https://doi.org/10.1002/mp.12256
28. Kim Y, Choi D, Lee KJ, Kang Y, Ahn JM, Lee E et al (2020) Ruling out rotator cuff tear in shoulder radiograph series using deep learning: redefining the role of conventional radiograph. Eur Radiol 30(5):2843–2852. https://doi.org/10.1007/s00330-019-06639-1
29. Lee K, Kim JY, Lee MH, Choi CH, Hwang JY (2021) Imbalanced loss-integrated deep-learning-based ultrasound image analysis for

diagnosis of rotator-cuff tear. Sensors 21(6):2214. https://doi.org/10.3390/s21062214
30. P.I. F. (2018) A tutorial on Bayesian optimization. ArXiv preprint. arXiv:1807.02811
31. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. 618–626
32. Simonyan K, Andrew Z (2014) Very deep convolutional networks for large-scale image recognition. ArXiv preprint. arXiv:1409.1556
33. He K, Zhang Z, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778
34. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV2016.
35. G. Huang ZL, L. Van Der Maaten and K. Q. Weinberger. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2017.
36. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 1251–1258
37. Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput 29(9):2352–2449. https://doi.org/10.1162/NECO_a_00990
38. Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ et al (2020) COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation. J Med Internet Res 22(6):e19569. https://doi.org/10.2196/19569
39. Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH (2019) Hyperparameter optimization for machine learning models based on Bayesian optimization. J Electron Sci Technol 17(1):26–40. https://doi.org/10.11989/JEST.1674-862X.80904120
40. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manage 45(4):427–437. https://doi.org/10.1016/j.ipm.2009.03.002
41. Medina G, Buckless CG, Thomasson E, Oh LS, Torriani M (2021) Deep learning method for segmentation of rotator cuff muscles on MR images. Skeletal Radiol 50(4):683–692. https://doi.org/10.1007/s00256-020-03599-2

Thao Thi Ho obtained a M.S. degree in nuclear and atomicphysics from Vietnam Academy of Science and Technology, in 2019. She iscurrently pursuing the Ph.D. degree in mechanical engineering at KyungpookNational University, Daegu, Republic of Korea. Her research interest includesmedical image processing, image registration, and segmentation based on deeplearning.

Geun-Tae Kim obtained a M.S. in physiology, from BusanNational University College of Medicine, Busan, Korea. He obtained a Ph.D. inInternal medicine, from Busan National University College of Medicine, Republicof Korea. He is currently a professor at Department of Internal Medicine,College of Medicine, Kosin University, Busan, Korea.

Taewoo Kim obtained a B.S. in Mechanical Engineering fromKyungpook National University, Korea. He is currently pursuing the M.S. degreein machine learning with quantitative computed tomography imaging-based metricsat School of Mechanical Engineering, Kyungpook National University, Daegu,Republic of Korea. His recent work is about alteration of airways and lung forCT images due to cement dust exposure.

Sanghun Choi obtained a M.S. in Mechanical Engineering,majoring in finite element-based two-phase flows, from Seoul National University,Korea. He obtained a Ph.D. in Mechanical Engineering, majoring in computationallung mechanics, from the University of Iowa, USA. He is currently an assistantprofessor at School of Mechanical Engineering, Kyungpook National University,Republic of Korea.

Eun-KeePark is a professor at Department of Medical Humanities and Social Medicine, College of Medicine, Kosin University. He received his Ph.D. degree in Environmental Toxicology from University ofSydney and his M.S. degree in Biotechnology, University of New South Wales. His research focuses on environmental and occupational health, evaluation ofenvironmental toxicants and multidisciplinary approaches to global health. Heis particularly interested in areas of pulmonary diseases and musculoskeletaldisorders.