

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

National Taiwan University

Master Thesis



以卷積神經網路實現棘上肌鈣化性肌腱炎之診斷

Diagnosis of Supraspinatus Calcific Tendinopathy using
Convolutional Neural Networks

邱霈欣

Pei-Hsin Chiu

指導教授：張書瑋 博士

Advisor: Shu-Wei Chang, Ph.D.

中華民國 109 年 7 月

July 2020

國立臺灣大學碩士學位論文 口試委員會審定書

以卷積神經網路實現棘上肌鈣化性肌腱炎之診斷
Diagnosis of Supraspinatus Calcific Tendinopathy
using Convolutional Neural Networks

本論文係 邱霈欣 君 (R07521219) 在國立臺灣大學土木工程學系
完成之碩士學位論文，於民國 109 年 7 月 29 日承下列考試委員
審查通過及口試及格，特此證明

口試委員：

張書瑋

(指導教授)

張書瑋

吳爵宏

吳爵宏

周佳靚

周佳靚

系主任

謝尚賢

謝尚賢

致謝

這份論文的完成雖然不算什麼大事，但是一路上卻要感謝許多師長與朋友所給予的幫助，首先要感謝的是我的指導教授張書瑋老師，謝謝他邀請了我一起參加導生宴而成為了進實驗室的契機，從我大四開始便給予我許多研究的機會，也謝謝他在這個題目上對我的支援與幫助，他在我接下來的求學路上所給予的種種建議與實質幫助也是讓我獲益良多，也要謝謝周佳靚教授在我剛開始學習機器學習時給了我很多的指導以及實習機會，還要謝謝一起完成這個研究的吳爵宏醫師，他給予剛開始做研究的我很多信心，也提供我許多研究上的建議以及臨床背景知識，三位老師對於我的指導與鼓勵都讓我銘記在心。

另外，我也想謝謝我的家人，謝謝我的爸爸總是讓我對自己的研究充滿自信，謝謝他包容我每次研究卡關時的壞心情，謝謝我的媽媽總是給予我追求夢想的空間，謝謝她給我經濟上的支援，讓我可以把握每次出國交流的機會，也要感謝我的弟弟，為我提供許多技術上的支援，讓我在入門機器學習這個領域時心裡很踏實，除了家人以外，我要感謝維翰學長、羽白、詩傑、諺霖以及所有給我技術指導與幫助的實驗室朋友們，沒有他們我的研究很難順利完成，也謝謝我的好友馨儀幫我一起記住我的研究的初衷，我希望未來能夠走得離這份初衷越來越近。

謝謝在我求學道路上幫助過我的所有師長以及同學和朋友們，雖然很難在這裡一一提及，但是大家對我不管是學術上、生活上亦或是心靈上的幫助都是造就我順利完成這份論文的重要的一環。最後，期許自己在完成碩士學位之餘，能夠繼續保持這份熱忱與信心，在求知的道路上更加精進。

摘要

以超音波檢查進行棘上肌鈣化性肌腱炎診斷時，常受限於其對操作者的經驗依賴性、放射科醫師人手不足以及城鄉之間民眾對於放射科服務可及性的差異，為了減緩以上之限制，本研究應用卷積神經網路建立人工智慧之機器學習模型，透過此模型建立電腦輔助系統以協助放射科醫師在超音波檢查中判斷長軸以及短軸的棘上肌鈣化性肌腱炎的存在。

本研究提出三個基於不同訓練集的機器學習模型，分別是僅基於長軸的長軸模型、僅基於短軸的短軸模型以及同時基於長、短軸的長短軸模型，以卷積神經網路之DenseNet121預訓練模型進行模型之建立與訓練，比較三者之間對於棘上肌鈣化性肌腱炎的判斷準確度以及其餘相關評估指標可知：同時基於長、短軸訓練之長短軸模型，其準確度為91.32%、靈敏度為87.89%、特異度為94.74%，是三種模型中最適合向放射科醫師提供超音波檢查中判斷長軸以及短軸的棘上肌鈣化性肌腱炎的存在之協助之模型。

關鍵字：棘上肌鈣化性肌腱炎；超音波；深度學習；機器學習；卷積神經網路；DenseNet121；醫療影像辨識

Abstract

Diagnosing supraspinatus calcific tendinopathy with ultrasound examination is usually limited by the dependence of ultrasound examination on experience of operators, the shortage of physicians, and the difference of the radiology service accessibility between urban and rural community. To eliminate these limitations, convolutional neural network was applied to build the machine learning models in this study. It was expected to build the computer aided system in order to assist physicians with diagnosing the existence of longitudinal view and transverse view of supraspinatus calcific tendinopathy in ultrasound examinations.

Three models trained on different training dataset were proposed in this study, which were the longitudinal model trained on only longitudinal view of supraspinatus calcific tendinopathy, the transverse model trained on only transverse view of supraspinatus calcific tendinopathy, and the longi-trans model trained on both longitudinal view and transverse view of supraspinatus calcific tendinopathy. These models were built and trained by DenseNet121, which is a pre-trained model in convolutional neural network. Compared the accuracy and other evaluation index of these three models against the longitudinal view and transverse view of supraspinatus calcific tendinopathy, it was found that the accuracy of the longi-trans model was 94.74%, which was the highest among the three models, and its sensitivity and specificity were 87.89% and 94.74% respectively. In conclusion, the longi-trans model is the most suitable model to provide physicians assistance of diagnosing the existence of longitudinal view and transverse view of supraspinatus calcific tendinopathy in ultrasound examinations.

Keywords: Supraspinatus calcific tendinopathy; Ultrasound; Deep learning; Machine learning; Convolutional Neural Network; DenseNet121; Medical images identification

目 錄



口試委員會審定書.....	i
致謝.....	ii
中文摘要.....	iii
英文摘要.....	iv
Chapter 1 Introduction.....	1
1.1 Background.....	1
1.2 Machine learning.....	4
1.2.1 Artificial intelligence.....	4
1.2.2 Machine learning.....	5
1.2.3 Deep learning.....	6
1.2.4 Convolutional neural networks.....	6
1.3 Statement of purpose.....	8
Chapter 2 Data description.....	9
2.1 Data collection.....	9
2.2 Data distribution.....	10
2.2.1 Longitudinal view of supraspinatus tendon.....	10
2.2.2 Transverse view of supraspinatus tendon.....	10
2.3 Data preprocessing.....	11
2.3.1 De-identification.....	11
2.3.2 Labeled ultrasound images.....	11
2.3.3 Data distribution.....	14
Chapter 3 Methodology.....	19

3.1 DenseNet-121.....	19
3.2 Data augmentation.....	20
3.2.1 Rotation range.....	21
3.2.2 Width shift range.....	22
3.2.3 Height shift range.....	22
3.2.4 Shear range.....	22
3.2.5 Zoom range.....	22
3.2.6 Channel shift range.....	23
3.2.7 Horizontal flip.....	23
3.2.8 Fill mode.....	23
3.2.9 Data flow.....	24
3.3 Dropout layer.....	25
3.4 Activation.....	25
3.5 Optimizer.....	26
3.5.1 Adam.....	26
3.5.2 Learning rate.....	27
3.6 Loss function.....	27
3.7 Class weight.....	28
3.8 Transfer learning.....	29
3.9 Heatmap.....	29
3.10 Evaluation index.....	31
3.10.1 Accuracy.....	32
3.10.2 Loss.....	32
3.10.3 Confusion matrix.....	33
3.10.4 Testing accuracy.....	34

3.10.5 Sensitivity.....	34
3.10.6 Specificity.....	35
3.10.7 Positive predictive value (PPV).....	35
3.10.8 Negative predictive value (NPV).....	35
3.10.9 False positive value (FPR).....	35
3.10.10 False negative value (FNR).....	36
3.10.11 Positive likelihood ratio (+LR).....	36
3.10.12 Negative likelihood ratio (-LR).....	37
3.10.13 Receiver operating characteristic curve (ROC curve).....	37
3.10.14 Area under the Curve (AUC).....	39
Chapter 4 Results.....	41
4.1 The longitudinal model.....	41
4.1.1 Evaluation index.....	41
4.1.1.1 Testing dataset with longitudinal view of supraspinatus tendon.....	41
4.1.1.2 Testing dataset with transverse view of supraspinatus tendon.....	44
4.1.1.3 Testing dataset with both longitudinal and transverse view of supraspinatus tendon.....	47
4.1.2 Heatmap.....	50
4.1.2.1 Testing dataset with longitudinal view of supraspinatus tendon.....	50
4.1.2.2 Testing dataset with transverse view of supraspinatus tendon.....	51
4.2 The transverse model.....	52

4.2.1 Evaluation index.....	52
4.2.1.1 Testing dataset with longitudinal view of supraspinatus tendon.....	52
4.2.1.2 Testing dataset with transverse view of supraspinatus tendon.....	55
4.2.1.3 Testing dataset with both longitudinal and transverse view of supraspinatus tendon.....	58
4.2.2 Heatmap.....	61
4.2.2.1 Testing dataset with longitudinal view of supraspinatus tendon.....	61
4.2.2.2 Testing dataset with transverse view of supraspinatus tendon.....	62
4.3 The longi-trans model.....	63
4.3.1 Evaluation index.....	63
4.3.1.1 Testing dataset with longitudinal view of supraspinatus tendon.....	63
4.3.1.2 Testing dataset with transverse view of supraspinatus tendon.....	66
4.3.1.3 Testing dataset with both longitudinal and transverse view of supraspinatus tendon.....	69
4.3.2 Heatmap.....	72
4.3.2.1 Testing dataset with longitudinal view of supraspinatus tendon.....	72
4.3.2.2 Testing dataset with transverse view of supraspinatus tendon.....	73

Chapter 5 Discussion.....	74
5.1 Evaluation index.....	74
5.2 Heatmap.....	80
Chapter 6 Conclusion.....	86
Reference.....	88



LIST OF FIGURES

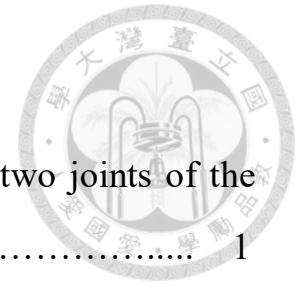


Figure 1-1: The humerus, clavicle, and scapula form the two joints of the shoulder.....	1
Figure 1-2: Schematic diagram of a rotator cuff.....	2
Figure 1-3: An architecture of Convolutional Neural Networks.....	8
Figure 2-1: An ultrasound images underwent de-identification.....	11
Figure 2-2: An example of ultrasound images with annotations.....	13
Figure 2-3: Ultrasound images of calcification cases and no calcification cases of different views of supraspinatus tendons.....	13
Figure 2-4: The monthly distribution of ultrasound images of different views of supraspinatus tendons.....	14
Figure 3-1: The architecture of four different types of DenseNet models.	20
Figure 3-2: “flow_from_directory” function in Keras.....	24
Figure 3-3: The performance of different choice of learning rate.....	27
Figure 3-4: Miscellaneous colormaps.....	31
Figure 3-5: ROC curve in different classification results.....	38
Figure 3-6: The complete parameters in “roc_curve” function.....	39
Figure 4-1: Learning curve of the longitudinal model.....	41
Figure 4-2: ROC curve of the longitudinal model testing against longitudinal view of supraspinatus tendon.....	44
Figure 4-3: ROC curve of the longitudinal model testing against transverse view of supraspinatus tendon.....	47
Figure 4-4: ROC curve of the longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon.....	50

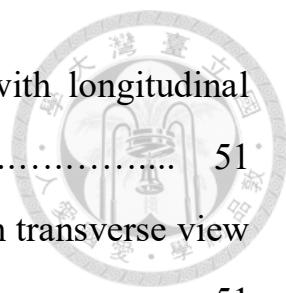


Figure 4-5: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the longitudinal model.....	51
Figure 4-6: Representative heatmaps of testing dataset with transverse view of supraspinatus tendon in the longitudinal model.....	51
Figure 4-7: Learning curve of the transverse model.....	52
Figure 4-8: ROC curve of the transverse model testing against longitudinal view of supraspinatus tendon.....	55
Figure 4-9: ROC curve of the transverse model testing against transverse view of supraspinatus tendon.....	58
Figure 4-10: ROC curve of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon.....	61
Figure 4-11: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the transverse model.....	62
Figure 4-12: Representative heatmaps of testing dataset with transverse view of supraspinatus tendon in the transverse model.....	62
Figure 4-13: Learning curve of the longi-trans model.....	63
Figure 4-14: ROC curve of the longi-trans model testing against longitudinal view of supraspinatus tendon.....	66
Figure 4-15: ROC curve of the longi-trans model testing against transverse view of supraspinatus tendon.....	69
Figure 4-16: ROC curve of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon.....	72
Figure 4-17: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the longi-trans model.....	73
Figure 4-18: Representative heatmaps of testing dataset with transverse view	

of supraspinatus tendon in the longi-trans model.....	73
Figure 5-1: The testing accuracy between three models against three different distribution of testing dataset.....	76
Figure 5-2: Sensitivity between three models against three different distribution of testing dataset.....	77
Figure 5-3: Specificity between three models against three different distribution of testing dataset.....	78
Figure 5-4: Heatmaps with color distributed on supraspinatus tendon....	80
Figure 5-5: Heatmaps with color distributed under supraspinatus tendon	81
Figure 5-6: Heatmaps with color distributed in greater tuberosity.....	82
Figure 5-7: Heatmaps of longitudinal model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.....	83
Figure 5-8: Heatmaps of transverse model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.....	84
Figure 5-9: Heatmaps of longi-trans model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.....	85

LIST OF TABLES

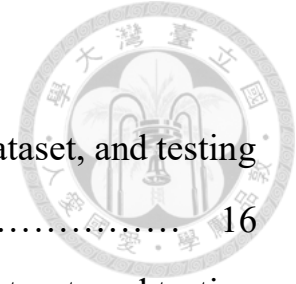


Table 2-1: The distribution of training dataset, validation dataset, and testing dataset in the longi-trans model.....	16
Table 2-2: The distribution of training dataset, validation dataset, and testing dataset in the longitudinal model.....	17
Table 2-3: The distribution of training dataset, validation dataset, and testing dataset in the transverse model.....	18
Table 3-1: Confusion matrix of a binary classification problem.....	34
Table 3-2: The clinical significance of +LR.....	37
Table 3-3: The meaning of AUC value.....	40
Table 4-1: Confusion matrix of the longitudinal model testing against longitudinal view of supraspinatus tendon.....	42
Table 4-2: Confusion matrix of the longitudinal model testing against transverse view of supraspinatus tendon.....	45
Table 4-3: Confusion matrix of the longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon.....	48
Table 4-4: Confusion matrix of the transverse model testing against longitudinal view of supraspinatus tendon.....	53
Table 4-5: Confusion matrix of the transverse model testing against transverse view of supraspinatus tendon.....	56
Table 4-6: Confusion matrix of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon.....	59
Table 4-7: Confusion matrix of the longi-trans model testing against longitudinal view of supraspinatus tendon.....	64
Table 4-8: Confusion matrix of the longi-trans model testing against	

transverse view of supraspinatus tendon.....	67
Table 4-9: Confusion matrix of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon.....	70
Table 5-1: Comparison between three models against three different distribution of testing dataset.....	79



Chapter 1 Introduction

1.1 Background

Shoulder is a ball-and-socket joint formed by the humerus, the clavicle, and the scapula as shown in Figure 1-1. These bones are wrapped by a layer of cartilage and form two main joints, acromioclavicular joint and glenohumeral joint. The latter is also known as the shoulder joint which is the most mobile joint in the body. The range of its motion comes from the rotator cuff. However, sports, repetitive movement, and manual labor will result in the injury of rotator cuff.

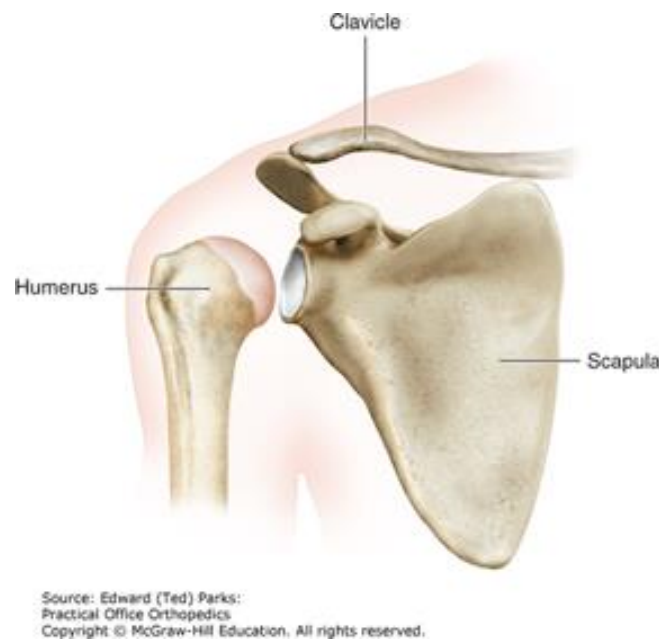


Figure 1-1: The humerus, clavicle, and scapula form the two joints of the shoulder.

Reprinted from *Practical Office Orthopedics*, by Edward (Ted) Parks, 2017. Copyright 2017 by McGraw-Hill Education.

Calcific tendinopathy of rotator cuff is the most common ultrasound (US)-diagnosed disorder [1-3]. It is caused by the abnormal deposition of calcium in tendon which is secondary to a local decrease in oxygen tension lead to fibrocartilaginous metaplasia and calcification. 80% of the cases of calcific tendinopathy occur in supraspinatus tendon which is higher than infraspinatus tendon and subscapularis tendon [4] as illustrated in Figure 1-2. It is recommended to observe the neutral, internal, and external rotation of rotator cuff to ensure the calcific tendinopathy occurs instead of other shoulder pathologies [5]. Furthermore, bilateral fashion is recommended as the conduct in imaging examinations in order to double check the existence of calcific tendinopathy [6].

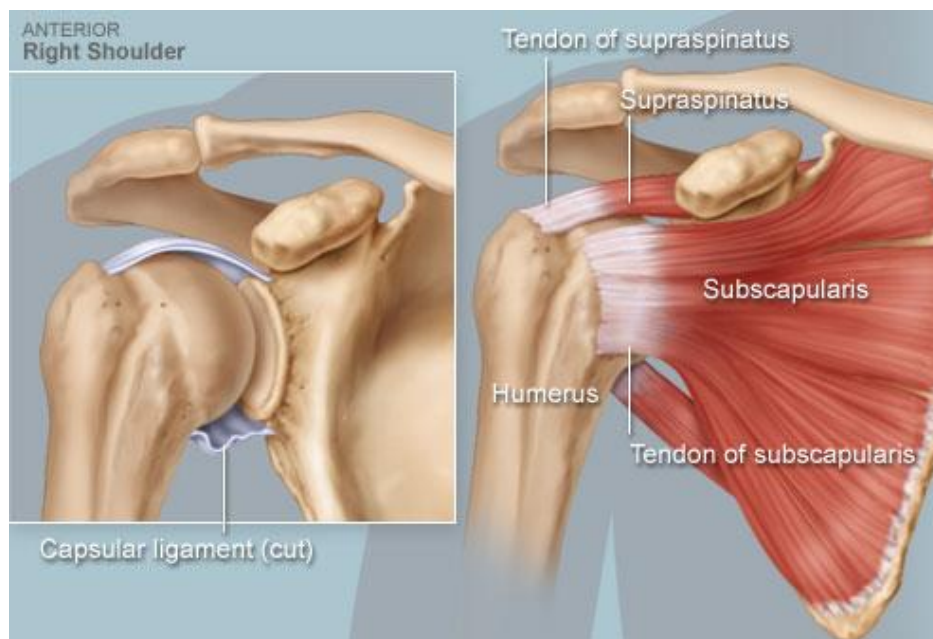
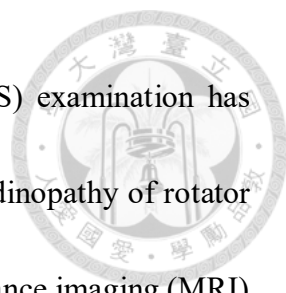
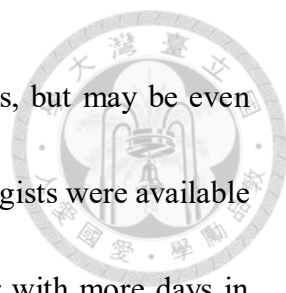


Figure 1-2: Schematic diagram of a rotator cuff. Reprinted from WebMD, LLC, 2014, from <https://www.webmd.com/pain-management/what-is-my-rotator-cuff#1>. Copyright 2014 by WebMD, LLC.



Among all the imaging examinations, B-mode ultrasound (US) examination has been considered as an excellent measure for visualizing calcific tendinopathy of rotator cuff [7, 8]. Its diagnosis accuracy was comparable to magnetic resonance imaging (MRI) [9-14] since the pain intensity in calcific tendinopathy of rotator cuff is related with the imaging features of US [15]. However, there are two weak points of US. First, US is known as an operator-dependent imaging modality that its diagnosis accuracy is highly related to the experience of the physicians [16, 17]. The significant training and practice determine if the diagnosis accuracy is in the clinically useful level. It was verified by O'Connor et al. [18] and Ohrndorf et al. [19] that a good interobserver agreement of US findings is reached among experienced physicians, but physicians reach a poor interobserver agreement when they are less experienced. The demarcation between experienced physicians and less experienced physicians can be inferred by the requirement of American College of Radiology (ACR) that physicians should perform 500 studies in a supervised setting within three years, and at least 100 exams per year in order to maintain the proficiency [20]. It was also reported by Hertzberg et al. [21] that 200 cases of US examinations was insufficient to develop a clinically useful level of competence in US. Second, the shortage of radiologists makes US examinations difficult to perform. Dallas-based U.S. Radiology Partners (USRP), a radiology management firm found that 45% of hospitals were understaffed in radiology in a survey of radiology




departments in 254 hospitals [22]. It not only happens in metropolis, but may be even common in rural community. Yawn et al. [23] pointed out that radiologists were available to hospitals in rural community for only 46% of the days each year with more days in larger hospital and fewer days in smaller hospitals. It caused a part of hospitals had only once-a-week, or even no on-site radiology services. The shortage of radiologists also increases the workload of the current radiologists. Soni et al. [24] found that an American radiologist worked around 50 hours per week which was higher than the basic workload formulated in the Labor law of America. As the workload increases, the pressure also increases. 45.8% of physicians in the survey done by Shanafelt et al. [25] reported that they had experienced burnout due to the workload and responsibilities. These phenomena may have impact on the diagnosis accuracy and the penetration rate of US.

1.2 Machine learning

To support the decision-making process of physicians and cut down their workload, computer-aided diagnosis was developed to provide objective opinions against the diagnosis of the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view for assisting physicians in this study.

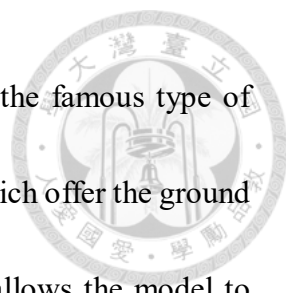
1.2.1 Artificial intelligence



Artificial intelligence (AI) is the intelligence displayed by machines manufactured by human. The concept of artificial intelligence was originated from Greek mythology, and its formal reasoning started from the first millennium BC. Until 1956, artificial intelligence was established as an academic discipline. AI was defined as “a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” by Andreas Kaplan and Michael Haenlein [26]. The application for artificial intelligence are distributed in intelligent control, robotics, automation, natural language processing, genetic programming, board game, medicine, and so on. Most of the current applications of artificial intelligence belong to weak AI, which is relative to strong AI. Weak AI is considered as intelligent apparently with no self-consciousness. Although weak AI is not as intelligent as strong AI, it still changes human's life by its learning ability.

1.2.2 Machine learning

Machine learning is a subset of artificial intelligence which is known for its experience-based learning mechanism. Mitchell [27] briefly defined it as a computer program that learn from experience E with respect to some class of tasks T and performance measure P , and improve with experience E when its performance at tasks in T , as measured by P . Different types of algorithm in machine learning derived from

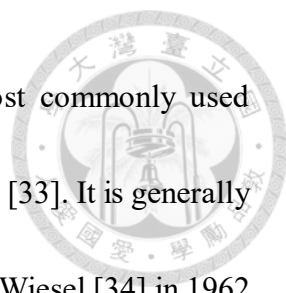


different types of purpose and data. Supervised learning as one of the famous type of algorithm in machine learning is highly dependent on labeled data which offer the ground truth to map an input to an output [28]. An appropriate algorithm allows the model to predict correctly unseen instances from the training data. In this study, supervised learning was chosen as the type of algorithm to develop the machine learning model against the diagnosis of the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view. We expect the model to predict the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view of unseen instances from the labeled training data in order to develop a machine learning model assist physicians.

1.2.3 Deep learning

Deep learning is a subset of machine learning which extract higher level features progressively from the input data with multiple layers [29]. It is based on artificial neural networks (ANN) with representation learning [30-32], which is a kind of computer system inspired from the biological neural networks composed from neurons, connections and weights, and propagation function. Convolutional neural network is the commonly based ANN in the most of deep learning models nowadays.

1.2.4 Convolutional neural network



Convolutional neural network abbreviated as CNN is the most commonly used method when dealing with images issues, especially in medicine field [33]. It is generally considered that the cat's visual cortex experiment done by Hubel and Wiesel [34] in 1962 is the initial idea of CNN. Until 1998, LeNet was proposed by LeCun et al. [35] which was the originator of CNN. However, AlexNet proposed by Krizhevsky et al. [36] in the 2012 ImageNet challenge is the one who really made CNN well-known. After AlexNet was proposed, there were also VGGNet proposed by Simonyan and Zisserman [37] in 2014, GoogLeNet proposed by Szegedy et al. [38] in 2014, ResNet proposed by He et al. [39] in 2015, and DenseNet proposed by Huang [40] in 2017. These CNN architectures are comprised of convolutional layers, pooling layers, and fully-connected layers. A CNN architecture is illustrated in Figure 1-3. Input data going through the CNN architecture is extracted highly representative features which is neither obvious nor easily handcrafted in US [41]. Therefore, in this study, we applied CNN as the architecture in the deep learning method to develop a machine learning model to assist physicians diagnosing the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view.

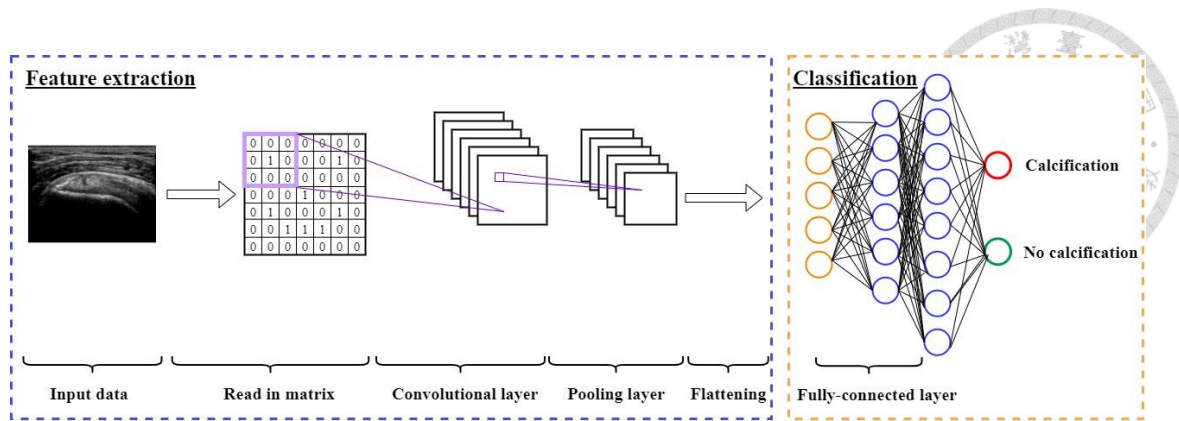


Figure 1-3: An architecture of Convolutional Neural Network.

1.3 Statement of purpose

Using convolutional neural network as the architecture to develop a machine learning model, our aim is to create a weak AI model based on the judgement of the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view in order to assist physicians. The purpose of this model is threefold: (1) Eliminate the differences of the diagnostic accuracy among different experienced levels of physicians. (2) Reduce the adequacy of specialists in radiology between urban and rural community. (3) Promote the effectiveness of US examinations and avoid physicians from overwork.

Chapter 2 Data description




2.1 Data collection

We obtained the institutional review board approval for this study, and were waived the requirement for informed consent for review of patient images and medical records. All ultrasonography examinations were performed by the 18 physiatrists whose experience are between 3 to 20 years. A variety of instruments including Hitachi, Siemens Acuson S2000, and Toshiba Canon Aplio 500 were used in the ultrasonography examinations which offered the ultrasound images in this study.

Ultrasound images in this study were collected from 1000 consecutive patients who underwent shoulder ultrasonography examinations. Longitudinal view and transverse view of supraspinatus tendons were included in the ultrasound images in this study. For longitudinal view of supraspinatus tendons, 7165 ultrasound images had been collected from January 2017 to June 2019. For transverse view of supraspinatus tendons, 5201 ultrasound images had been collected from January 2017 to December 2018.

In this study, we used examination-level labels rather than image-level labels to construct the calcification diagnosis model. Images collected from the same ultrasonography examination shared the same label which represents the images with or without supraspinatus calcific tendinopathy. Two conditions were used to be the grounds



of diagnosis, including the presence of hyperechoic lesion and the acoustics shadow in the supraspinatus tendon. An image was diagnosed as “with supraspinatus calcific tendinopathy” when it met both conditions. Other findings such as supraspinatus tendon tear, subdeltoid bursitis or tendinitis were not included in the annotation. During annotation, 2 physiatrists labeled ultrasound images as with or without supraspinatus calcific tendinopathy independently. If antilogies existed, consensus was reached by discussion.

2.2 Data distribution

2.2.1 Longitudinal view of supraspinatus tendons

7165 ultrasound images of longitudinal view of supraspinatus tendons collected from January 2017 to June 2019 included with supraspinatus calcific tendinopathy cases and without supraspinatus calcific tendinopathy cases. In the following paragraphs, calcification cases and no calcification cases are used to represent each of them.

The amount of calcification cases and no calcification cases of longitudinal view are 2222 and 4943 separately.

2.2.2 Transverse view of supraspinatus tendons

5201 ultrasound images of transverse view of supraspinatus tendons collected from January 2017 to December 2018 included calcification cases and no calcification cases. The amount of calcification cases and no calcification cases of transverse view are 1178 and 4023 separately.

2.3 Data preprocessing

2.3.1 De-identification

In order to protect the security of patients' personal information, de-identification was implemented. Information including patients' names, names of hospitals and ultrasound machines was all removed during the data preprocessing process. An ultrasound images underwent de-identification is shown as Figure 2-1.

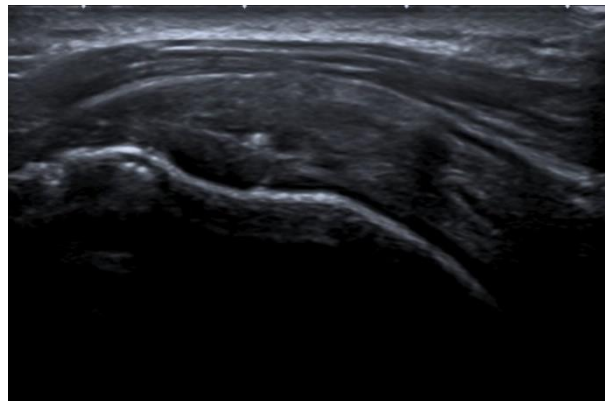



Figure 2-1: An ultrasound images underwent de-identification.

2.3.2 Labeled ultrasound images



Several annotations were observed on part of the collected ultrasound images as shown in Figure 2-2. They stand for the position of calcification occurrence labeled by physiatrists. These labeled ultrasound images account for 63.9% in calcification cases and 23.4% in no calcification cases for longitudinal view of supraspinatus tendons, and 61.7% in calcification cases and 29.8% in no calcification cases for transverse view of supraspinatus tendons. In this study, labeled ultrasound images were excluded from the dataset since the original ultrasound images are without any labels. Figure 2-3 shows the ultrasound images of calcification cases and no calcification cases of different views of supraspinatus tendons. The amount of calcification cases and no calcification cases of longitudinal view of supraspinatus tendons without labels are 780 and 3784, and those of transverse view of supraspinatus tendons without labels are 451 and 2822 separately. The monthly distribution of ultrasound images of longitudinal view and transverse view of supraspinatus tendons are shown as Figure 2-4. Calcification cases and no calcification cases of longitudinal view of supraspinatus tendons without labels were doubled checked by physiatrists to ensure the correctness of the label.

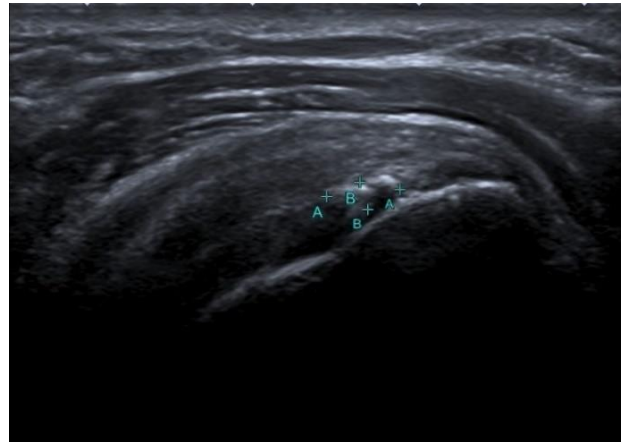


Figure 2-2: An example of ultrasound images with annotations.

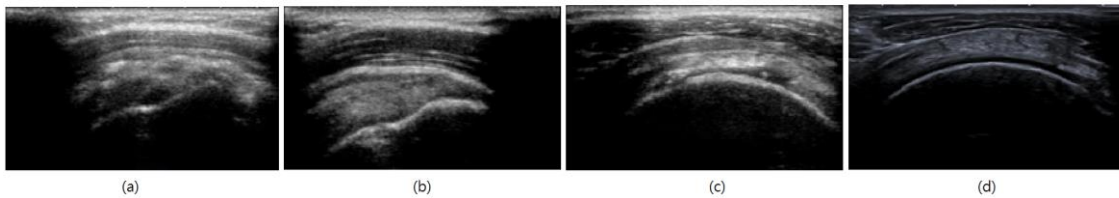
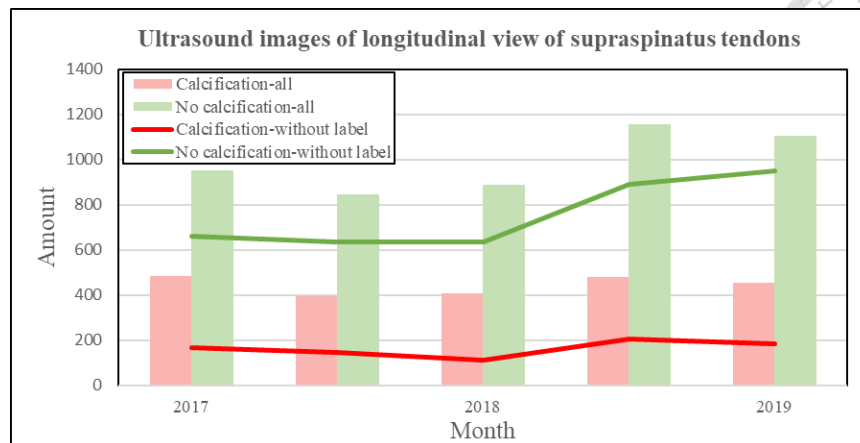
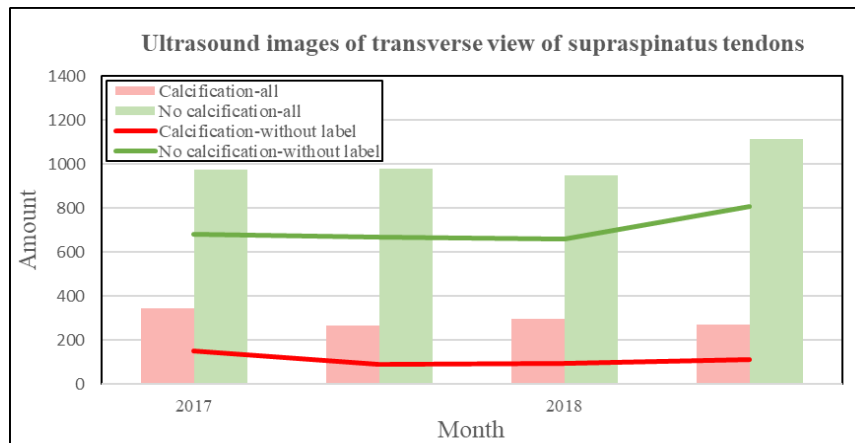


Figure 2-3: Ultrasound images of calcification cases and no calcification cases of different views of supraspinatus tendons. (a) Ultrasound images of calcification cases of longitudinal view of supraspinatus tendons. (b) Ultrasound images of no calcification cases of longitudinal view of supraspinatus tendons. (c) Ultrasound images of calcification cases of transverse view of supraspinatus tendons. (d) Ultrasound images of no calcification cases of transverse view of supraspinatus tendons.



(a)

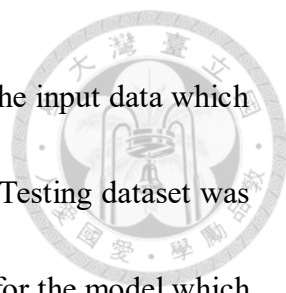


(b)

Figure 2-4: The monthly distribution of ultrasound images of different views of supraspinatus tendons. (a) The monthly distribution of ultrasound images of longitudinal views of supraspinatus tendons. (b) The monthly distribution of ultrasound images of transverse views of supraspinatus tendons.

2.3.3 Dataset distribution

We distributed the ultrasound images into training dataset, validation dataset, and testing dataset. Training dataset and validation dataset were used in the training process.



The former offered the input data for training, and the latter offered the input data which validated the training process in order to fix the learning procedure. Testing dataset was used in predicted process. Data in testing dataset are brand new data for the model which stands for on-site ultrasound images. The distribution of these datasets are shown in Table 2-1, 2-2, and 2-3. We designed three models with different distributions of datasets, which were the longitudinal model, the transverse model, and the longi-trans model. For the longi-trans model, there were 600 and 50 ultrasound images without labels of calcification cases and no calcification cases of longitudinal view of supraspinatus tendons, and 360 and 31 ultrasound images without labels of calcification cases and no calcification cases of transverse view of supraspinatus tendons in their training datasets and validation datasets separately. For the longitudinal model, only 600 and 50 ultrasound images without labels of calcification cases and no calcification cases of longitudinal view of supraspinatus tendons were included in its training dataset and validation dataset. For the transverse model, only 360 and 31 ultrasound images without labels of calcification cases and no calcification cases of transverse view of supraspinatus tendons were included in its training dataset and validation dataset. All the three models tested on the dataset with longitudinal view of supraspinatus tendons only, with transverse view of supraspinatus tendons only, and with both view of supraspinatus tendons separately.

Table 2-1: The distribution of training dataset, validation dataset, and testing dataset in the longi-trans model. Testing dataset(a) with longitudinal view of supraspinatus tendon. (b) with transverse view of supraspinatus tendon. (c) with both longitudinal and transverse view of supraspinatus tendon.

	Training data	Validation data	Testing data(a)	Testing data(b)	Testing data(c)
All	1920	162	260	120	380
Calcification cases of longitudinal view of supraspinatus tendons	600	50	130	0	130
No calcification of longitudinal view of supraspinatus tendons	600	50	130	0	130
Calcification cases of transverse view of supraspinatus tendons	360	31	0	60	60
No calcification cases of transverse view of supraspinatus tendons	360	31	0	60	60

Table 2-2: The distribution of training dataset, validation dataset, and testing dataset in the longitudinal model. Testing dataset(a) with longitudinal view of supraspinatus tendon. (b) with transverse view of supraspinatus tendon. (c) with both longitudinal and transverse view of supraspinatus tendon.

	Training data	Validation data	Testing data(a)	Testing data(b)	Testing data(c)
All	1200	100	260	120	380
Calcification cases of longitudinal view of supraspinatus tendons	600	50	130	0	130
No calcification of longitudinal view of supraspinatus tendons	600	50	130	0	130
Calcification cases of transverse view of supraspinatus tendons	0	0	0	60	60
No calcification cases of transverse view of supraspinatus tendons	0	0	0	60	60

Table 2-3: The distribution of training dataset, validation dataset, and testing dataset in the transverse model. Testing dataset(a) with longitudinal view of supraspinatus tendon. (b) with transverse view of supraspinatus tendon. (c) with both longitudinal and transverse view of supraspinatus tendon

	Training data	Validation data	Testing data(a)	Testing data(b)	Testing data(c)
All	720	62	260	120	380
Calcification cases of longitudinal view of supraspinatus tendons	0	0	130	0	130
No calcification of longitudinal view of supraspinatus tendons	0	0	130	0	130
Calcification cases of transverse view of supraspinatus tendons	360	31	0	60	60
No calcification cases of transverse view of supraspinatus tendons	360	31	0	60	60

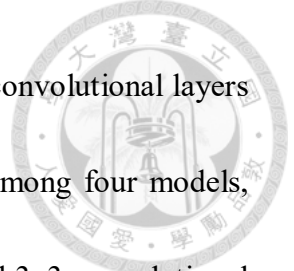
Chapter 3 Methodology



The longitudinal model, the transverse model, and the longi-trans model shared the same parameters illustrated in this chapter. All the model was developed with Tensorflow and Keras in Python language.

3.1 DenseNet-121

DenseNet-121 was the pre-trained network used in this study. Its weight is trained on ImageNet, a pre-trained dataset with more than 14 million images over 20,000 classes. DenseNet-121, as its name suggested, is a DenseNet model with 121 layers. 4 classical types of DenseNet were published, which are DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264. Figure 3-1 shows the architectures for different DenseNet models. These four models went with a 7x7 convolutional layer with stride equaled to 2 which had two kernels first, and were followed by a 3x3 Max Pooling layer with stride equaled to 2. Dense block layers and transition layers were followed after the two layers with same configuration among the four models until the third group of dense block layer and transition layer. There were 24 groups of 1x1 convolutional layers and 3x3 convolutional layers in the dense block (3) layer of DenseNet-121. For DenseNet-169, 32 groups of 1x1 convolutional layers and 3x3 convolutional layers composed the dense block (3) layer. The dense block (3) layer of DenseNet-201 and DenseNet-264 were



composed of 48 and 64 groups of 1x1 convolutional layers and 3x3 convolutional layers separately. Continued with same configuration of transition layer among four models, there were 16, 32, 32, and 48 groups of 1x1 convolutional layers and 3x3 convolutional layers in the dense block (4) layer of four models separately. A 7x7 global average pooling layer was connected with the aforementioned structure in the four models, and a fully connected layer followed behind.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 3-1: The architecture of four different types of DenseNet models [40].

In this study, the top fully connected layer of DenseNet-121 was replaced with 2 classes for the transverse model, the longitudinal model, and 4 classes for the longi-trans model as the dense layers.

3.2 Data augmentation

Collecting medical images was not as easy as collecting general images. The

prevalence rate of specific disease might be a main cause. Even if one suffers from the disease, the cost gap between medical images and general images would be another factor that leads to the scarcity of medical images. The adequacy of dataset is highly correlated with the performance of machine learning model. An insufficient dataset is hard to build an effective model because it is not able to learn completely with few data.

To reduce the impact of the shortage of our ultrasound images, data augmentation was used to diversify the input data. Data augmentation function under keras is called “ImageDataGenerator”. There are 22 parameters including the changes of rotation angles, image sizes, and so on. Here we only discuss the parameters used in our model. Rotation mentioned in 3.2.1 was used to simulate the difference between the size and position of supraspinatus tendon of patients. Moreover, width shift, height shift, shear, zoom, and channel shift mentioned from 3.2.2 to 3.2.6 were used to simulate the difference made by the personal error of physicians, and channel shift in 3.2.7 was used to simulate the color difference between multiple brands of ultrasound machine. Horizontal flip mentioned in 3.2.8 was used instead of vertical flip in this study in order to simulate difference in direction between the right and left shoulders.

3.2.1 Rotation range

Rotation range is the degree range of random rotation which is prescribed as an

integer. It is set as 40 that the rotation degree is random from 0 to 40 degree in each batch during data augmentation process in this study.



3.2.2 Width shift range

Width shift range is the range of random width shift which is prescribed as an integer, a float, or a 1-D array-like. It is set as 0.2 in this study that the width is shifted randomly from 0 to 0.2 level in each batch during data augmentation process in this study.

3.2.3 Height shift range

Height shift range is the range of random height shift which is prescribed as an integer, a float, or a 1-D array-like. It is set as 0.2 that the height is shifted randomly from 0 to 0.2 level in each batch during data augmentation process in this study.

3.2.4 Shear range

Shear range is the range of shear transform which is prescribed as a float. It fixes one axis (x axis or y axis), and shifts another axis (y axis or x axis) in proportion. It is set as 0.2 that the height is shifted randomly from 0 to 0.2 level in each batch during data augmentation process in this study.

3.2.5 Zoom range

Zoom range is the range of random zoom which is prescribed as a float. It is set as 0.2 that the image which undergoes zoom is zoomed randomly from 0 to 0.2 level in both length and width simultaneously in each batch during data augmentation process in this study.

3.2.6 Channel shift range

By shifting the value of channel, an image changes color after its channel is shifted. It is prescribed as a float. It is set as 0.2 that the color of image is changed because of the channel shifted randomly from 0 to 0.2 level in each batch during data augmentation process in this study. While enhancing the images, it keeps a certain degree of similarity of original images simultaneously.

3.2.7 Horizontal flip

The value of horizontal flip is prescribed as Boolean. In this study, an image is flipped randomly and horizontally in each batch during data augmentation process by setting the value of horizontal flip as true.

3.2.8 Fill mode

When deformation happened, there may have some missing regions. To fill these regions, there were four different kinds of fill modes in the “ImageDataGenerator”

function. Among these four kinds of fill modes, we used nearest as the fill mode during the image deformation in this study.



3.2.9 Data flow

Augmentative data were flowed into the model by the “flow_from_directory” function in Keras. The input of the function in this study is shown in Figure 3-2. The target size was defaulted as (256, 256) originally, but it was adjusted to (224, 224) because of the requirement of ImageNet in this study. The mode of color was set as the default of “rgb” in this study. The corresponding labels of the input data were got from the names of directories automatically since we set the parameter, classes, as none. There were five choices in the parameter, “class_mode”, which were categorical, binary, sparse, input, and none. In this study, categorical was chosen to return 2D one-hot encoded labels. We set batch size as 8 and turned on the shuffle function to shuffle data. The parameter, “interpolation” was used to adjust the size of input images when their sizes were different from the target size. There were three types of interpolation, including nearest, bilinear, and bicubic. In this study, bicubic was chose as the type of interpolation.

```
flow_from_directory(directory, target_size=(224, 224),  
class_mode='categorical', batch_size=8, shuffle=True,  
interpolation='bicubic')
```

Figure 3-2: “flow_from_directory” function in Keras.

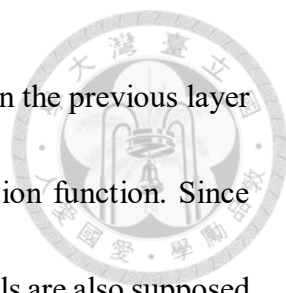
3.3 Dropout Layer



Overfitting is a common phenomenon happened when a machine learning model is lack of training data or full of parameters. It is observable when training loss is lower than validation loss and training accuracy is higher than validation accuracy simultaneously. An overfitting model over fits the features of a specific group of data which would lead to a bad performance when facing brand new data, although its accuracy seems to be high enough. Before Hinton [42] proposed the concept of dropout in 2012, training multiple models was one way to deal with overfitting which derived the problem of time-consuming. The power of dropout was proved in the 2012 ImageNet large scale visual recognition challenge with the AlexNet model proposed by Alex et al [36].

In this study, we avoided overfitting by adding a dropout layer with dropout rate 0.5. This dropout layer would delete half of the hidden neurons in the network randomly, and kept the amount of the input and output neurons simultaneously in each batch. The surviving neurons would continually update the parameters with the optimizer. By abandoning half of the hidden neurons in random, we were able to prevent the model from over fits the features in a specific group of training data.

3.4 Activation



A linear relation between input and output is existing that input in the previous layer would become the output in the next layer when there is no activation function. Since most of the issues can be looked as non-linear, machine learning models are also supposed to use non-linear equations to deal with these issues.

In this study, softmax function was taken as the activation function in the model. It can be looked as the generalization of sigmoid, but more useful in multiple classification issues. Softmax function in the model not only cut off the linear relation between input and output, but also converted the value of prediction into a numerical value between zero to one. It simplified the way we presented the predicted results of the model that the output of the model would be a probability of supraspinatus tendons with supraspinatus calcific tendinopathy.

3.5 Optimizer

3.5.1 Adam

An appropriate optimizer is able to minimize loss function and improve the training process in the model. In this study, we took Adam as the optimizer rather than commonly used stochastic gradient descent(SGD) for two reasons. First, its low memory usage and good computing performance was beneficial to our training process. Second, it is good at handling noise samples and sparse gradient. When the input dataset is sparse, Adam as an



optimizer in the model would adjust the learning rate automatically in order to get the best performance.

3.5.2 Learning rate

Learning rate is related with the performance of gradient descent. A too large learning rate is invalid for the training and the training process would be unable to converge which is shown as Figure 3-3(c). On the other hand, a too small learning rate also leads to an inefficient training process because there are too much updates before reaching the minimum point which is shown in Figure 3-3(a). Figure 3-3(b) is an appropriate learning rate which reaches the minimum point with a swift and optimal way. The learning rate we used in this study was set as 10^{-5} .

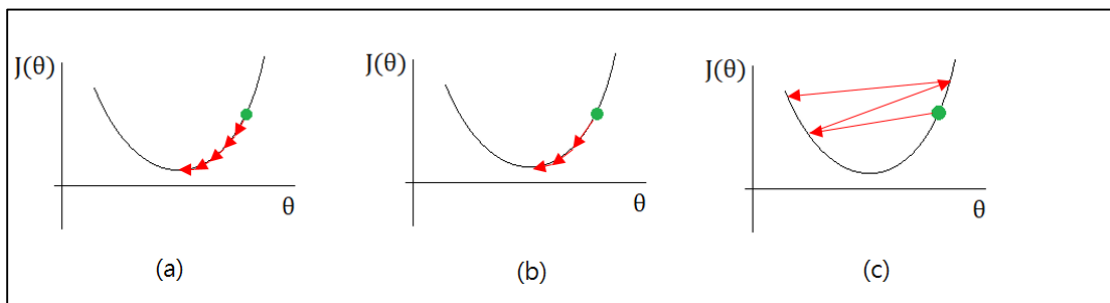
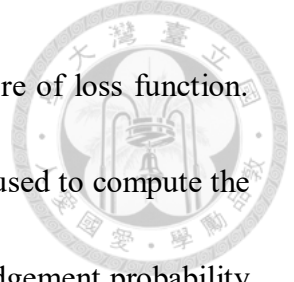


Figure 3-3: The performance of different choice of learning rate. (a) A too small learning rate (b) An appropriate learning rate (c) A too large learning rate.

3.6 Loss function



Residual between the ground truth and the prediction is the core of loss function. The smaller the residual is, the better the model is. Cross-entropy is used to compute the loss in the model instead of error rate because it is able to tell the judgement probability of true classes, which is the most important features in the learning process. There are nine types of loss function offered by Keras. ‘categorical_crossentropy’ was chose as the loss function in this study. The equation of ‘categorical_crossentropy’ is shown as (3.1). The parameter, ‘n’ stands for the number of samples, and ‘m’ stands for the number of categories. It is commonly used when softmax function was set in the dense layer.

$$\text{loss} = - \sum_{i=1}^n \hat{y}_{im} \log y_{im} \quad (3.1)$$

3.7 Class weight

As mentioned in 3.2, collecting medical images was not as easily as collecting general images. What’s more, imbalanced data usually occurs in medical images. Imbalanced data, as its name suggested, is the imbalance among each classes of data. In Keras, the function, “class_weight” is used to deal with it. This function is able to adjust the weights of data in an auto or a given way. For those classes with plentiful data, their weights are light, and for those with rare data, their weights become heavy after using the function, “class_weight” in Keras. In this study, data between different classes in the

longi-trans model were imbalanced data. Therefore, the function, “class_weight” was set as “auto” in the longi-trans model.

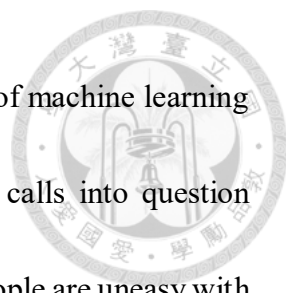


3.8 Transfer learning Freeze layer

Transfer learning is a field of research of machine learning which focuses on the reuse and transformation between the information of pre-learned tasks and new tasks [43]. It is an effective way to improve the learning process by learning new tasks based on some basic knowledge and cut down the training time. In this study, we took ImageNet as the pre-trained dataset to implement transfer learning. The top of fully connected layer in ImageNet was abandoned in order to connect the appropriate fully connected layer in this study. The first and the second layers were also frozen in order to make up the scarcity of the data.

3.9 Heatmap

Heatmap is a kind of data visualization which reveal data with colors in two dimensions. It replaces data with larger value with black squares, and smaller values with lighter squares. Heatmap was first applied by Loua (1873) [44] in sociological statistics, but widely used in biological sciences nowadays. Machine learning is another growing application field. The flourish development of machine learning not only brings up the



heavy application of machine learning in all walks of life, the doubt of machine learning follows behind. The decision-making process of machine learning calls into question because of the characteristic of machine learning called black box. People are uneasy with the black box in machine learning when it is related to their money or life. To get human's approval, heatmap is used in machine learning to visualize what the model "saw" in the input images in this study.

Gradient-weighted Class Activation Mapping(Grad-CAM), which was proposed by Selvaraju et al. [45] was used to plot heatmap in this study. It is derived from Class Activation Mapping(CAM), which was proposed by Zhou et al. [46]. The significant difference between Grad-CAM and CAM is that global average pooling is not necessary in the former by using back propagation to calculate weights. Weights are multiplied with feature maps in order to get the importance of regions in the last convolutional layer of CNN because the last convolutional layer is believed to have the most abundant information between detailed spatial information and high-level semantics.

Before calculating heatmap, images were loaded as PIL images and then transformed to n-dimensional array (ndarray) in Numpy. Each ndarray was added one batch size in the first axis. Grad-CAM was compiled after finishing image processing. We first got the prediction of images by "predict" function in Model class. Second, we calculated the gradients of the prediction in the last convolution layer, which was

“conv5_block16_2_conv” in DesneNet-121. Next, the sum of gradients of each feature map and the last convolution layer of the model were got by “function” function in the backend of Keras. The importance of certain regions in the feature map for the classification was got from the product of feature map and weights.

After calculating the value of heatmap, we plotted heatmap with “matplotlib” library. The original images were plotted with transparency as 0.6, and the heatmaps were covered on them with transparency as 0.4. In the choice of color of heatmap, Jet was chosen among other rules in miscellaneous colormaps shown as Figure 3-4.

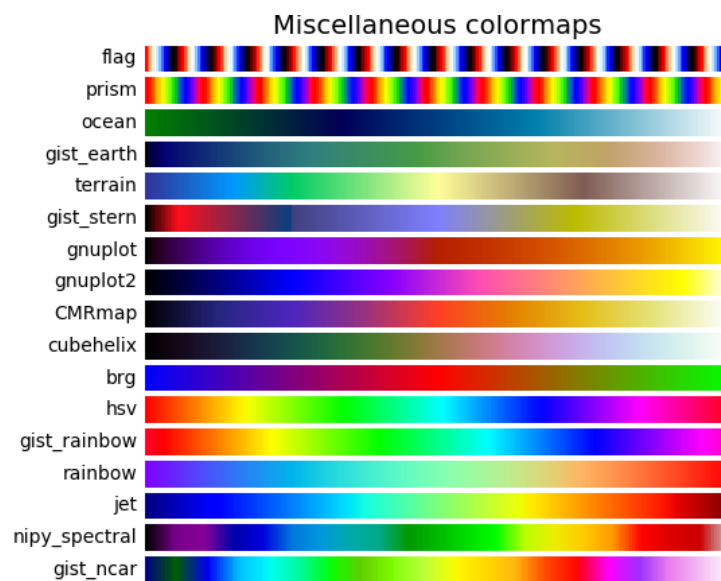


Figure 3-4: Miscellaneous colormaps.

3.10 Evaluation index

In this study, classification metrics was used as the evaluation index to assess the effectiveness of the model judging if supraspinatus tendons are with or without

supraspinatus calcific tendinopathy. The following 14 index were used to evaluate the model.



3.10.1 Accuracy

Accuracy is the rate of correct predictions including true positive and true negative among all cases. In this study, training accuracy, and validation accuracy were included in the metric, accuracy. Training accuracy was the accuracy during training process. It showed the performance while the model was trained on training dataset. In each epoch, validation accuracy was calculated in the validation process after training process. It reflected the performance while the model was validated on validation dataset. Training accuracy was not trustworthy as the accuracy of the model because a good training accuracy could even happen in a poorly trained model. On the other hand, validation accuracy was more credible as the representative of the performance of the model. Therefore, in this study, validation accuracy represents for the accuracy of the model.

3.10.2 Loss

Loss including training loss and validation loss is used to evaluate the performance of the model. It is calculated by the loss function we set in the model. When training loss is higher than validation loss, the model is under fitting which shows that the model is

still in its learning process and it needs more epoch and time to learn all features from data. When training loss is lower than validation loss, the model is over fitting which shows that it is fitting well in training data rather than validation data. It also shows that the model is unable to deal with unseen data. A perfect model exists when training loss is roughly equal to validation loss. Therefore, the relation between training loss and validation loss could be an evaluation index for the model.

3.10.3 Confusion Matrix

Confusion matrix in a binary classification problem is shown in Table 3-1. It shows the validity of the model. Actual classes are the ground truth, and predicted classes are the outcome predicted by the model. Positive stands for the supraspinatus tendons with supraspinatus calcific tendinopathy including longitudinal view and transverse view. Negative stands for the supraspinatus tendons without supraspinatus calcific tendinopathy including longitudinal view and transverse view in this study. True positive(TP) is the number of positive cases which are correctly predicted as positive. False positive(FP) is the number of negative cases which are incorrectly predicted as positive cases. False negative(FN) is the number of positive cases that are incorrectly predicated as negative cases, and true negative(TN) is the number of negative cases that are correctly predicted as negative cases.



Table 3-1: Confusion matrix of a binary classification problem.

There are eight evaluation indexes derived from confusion matrix which are sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV), false positive value(Fpr), false negative value(Fnr), positive likelihood ratio(+LR), and

		Actual classes	
		Positive	Negative
Predicted classes	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

negative likelihood ratio(-LR). These evaluation index used frequently in both machine learning field and medical clinic.

3.10.4 Testing accuracy

Testing accuracy is the rate of correct predictions including true positive and true negative among total number of cases. The formula of accuracy:

$$\text{Testing accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}} \quad (3.2)$$

3.10.5 Sensitivity

Sensitivity, also called true positive rate(TPR) or recall, is the rate of actual positive

that is correctly predicted as positive. The formula of sensitivity:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3.3)$$



3.10.6 Specificity

Specificity, also called true negative rate(TNR), is the rate of actual negative that is correctly predicted as negative. The formula of specificity:

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} \quad (3.4)$$

3.10.7 Positive predictive value (PPV)

Positive predictive value(PPV), also called the positive predictive agreement, stands for the rate of positive prediction among actual positive cases. The formula of PPV:

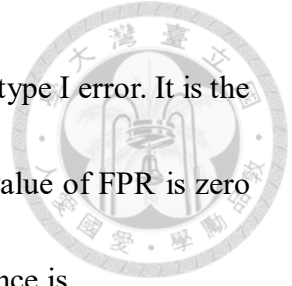
$$\text{PPV} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3.5)$$

3.10.8 Negative predictive value (NPV)

Negative predictive value(NPV), also called the negative predictive agreement, stands for the rate of negative prediction among actual negative cases. The formula of NPV:

$$\text{NPV} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}} \quad (3.6)$$

3.10.9 False positive value (FPR)



False positive value(FPR), commonly called as false alarm, is a type I error. It is the rate when a negative case is predicted as a positive case. The ideal value of FPR is zero because the lower the rate of misdiagnosis is, the better the performance is.

The formula of FPR:

$$FPR = \frac{False\ Positive}{False\ Positive + True\ Negative} = 1 - Specificity \quad (3.7)$$

3.10.10 False negative value (FNR)

False negative value(FNR), is a type II error. It is the rate when a positive case is predicted as a negative case. The ideal value of FNR is zero because the lower the rate of missed diagnosis is, the better the performance is. The formula of FNR:

$$FNR = \frac{False\ Negative}{True\ Positive + False\ Negative} = 1 - Sensitivity \quad (3.8)$$

3.10.11 Positive likelihood ratio (+LR)

Positive likelihood ratio(+LR) is the time of the probability of predicted positive cases than that of predicted negative cases. The higher the +LR is, the greater the probability of positive case is. A low +LR is according to negative cases. The formula of +LR:

$$+LR = \frac{(True\ Positive)/(True\ Positive + False\ Negative)}{(False\ Postive)/(False\ Postive + True\ Negative)} \quad (3.9)$$

The clinical significance of +LR is shown in Table 3-2.



Table 3-2: The clinical significance of +LR.

Positive likelihood ratio(+LR)	Clinical significance
>10	Strong evidence to rule in disease
5-10	Moderate evidence to rule in disease
2-5	Weak evidence to rule in disease
0.5-2.0	No significant change in the likelihood
0.2-0.5	Weak evidence to rule out disease
0.1-0.2	Moderate evidence to rule out disease
<0.1	Strong evidence to rule out disease

3.10.12 Negative likelihood ratio(-LR)

Negative likelihood ratio(-LR) is the time of the probability of predicted negative cases than that of predicted positive cases. A low -LR is according to a valued diagnosis.

The formula of -LR:

$$-LR = \frac{(False\ Negative)/(True\ Positive + False\ Negative)}{(True\ Negative)/(False\ Positive + True\ Negative)} \quad (3.10)$$

3.10.13 Receiver operating characteristic curve (ROC curve)

Receiver operating characteristic curve was first used to detect enemy vehicles with signal detection theory during World War II. It is a graphical plot that analysis the ability

of a binary-classifier with changed threshold. It has been widely used in meteorology, medicine, and radiology for decades. Recently, its application in machine learning and data mining has increased.



ROC curve is highly correlated with confusion matrix and its derived evaluation index mentioned above. In a ROC space, FPR is defined as the x axis, and TPR is defined as the y axis. As the threshold changes, the corresponding FPR and TPR are calculated and plotted as coordinated points. These coordinated points are connected as a ROC curve shown in Figure 3-5. The red diagonal in Figure 3-5, also called line of no-discrimination, stands for the random prediction with the accuracy of 50%. When a ROC curve is under the diagonal as Figure 3-5(a), it is treated as a bad classifier since its accuracy is even lower than the accuracy of random prediction. On the other hand, a good classifier looks like the ROC curve in Figure 3-5(c). The more the ROC curve is to upper left, the better the classifier is.

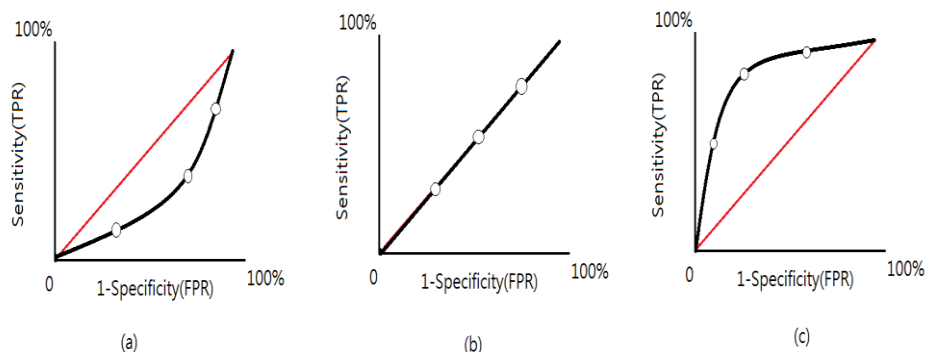
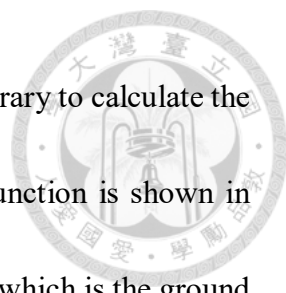


Figure 3-5: ROC curve in different classification results. (a)A bad classifier (b)A random classifier with accuracy of 50% (c)A good classifier.



In this study, we used the “roc_curve” function in scikit-learn library to calculate the ROC curve of the model. The complete input of the “roc_curve” function is shown in Figure 3-6. The parameter, “y_true” stands for the true binary labels, which is the ground truth in this study. The parameter, “y_score” stands for the target scores. It was considered as the probability of positive cases in this study. The parameters, “pos_label”, which stands for the label of positive class, “sample_weight”, which stands for sample weights, and “drop_intermediate”, which is used to create lighter ROC curves by dropping some suboptimal thresholds that would not appear on a plotted ROC curve, were kept in default values in this study. That is, to plot the ROC curve in this study, “y_true”, “y_score”, and “drop_intermediate” are used in the “roc_curve” function in scikit-learn library. The output of the “roc_curve” function is FPR, TPR, and thresholds. These output parameters are used to plot the ROC curve and calculate the area under curve.

```
sklearn.metrics.roc_curve(y_true, y_score, *, pos_label=None, sample_weight=None, drop_intermediate=True)
```

Figure 3-6: The complete parameters in “roc_curve” function.

3.10.14 Area under the Curve (AUC)

Area under the curve of ROC curve (AUC) is used to digitize the difference of different ROC curves since it is less accurate to compare different models with only ROC curves. There are three features of AUC. First, AUC must be between zero to one since

the ROC curve is a 1x1 grid. Second, calculation of AUC is based on the hypothesis that those cases with probability beyond thresholds are predicted as positive classes, and the others are negative classes. Third, the meaning of AUC value is shown in Table 3-3. AUC values which are higher than 0.5 are in different levels of discrimination. Those classifiers with AUC lower than 0.5 are also useful to do classify if they are used in opposite way. Only when AUC equals to 0.5, it has no contributions to classification. The higher the absolute value of AUC is, the better its discrimination is.

Table 3-3: The meaning of AUC value.

AUC	Meaning
0.9-1	Outstanding discrimination
0.8-0.9	Excellent discrimination
0.7-0.8	Acceptable discrimination
0.5	No discrimination
<0.5	Worse than random prediction

In this study, we used the “auc” function in scikit-learn to calculate the AUC of ROC. The complete input of the “auc” function is shown in (3.11). The parameters, x and y stand for the x coordinates and y coordinates separately. The output of the “auc” function is the AUC value of the ROC curve.

$$\text{sklearn.metrics.}\mathbf{auc}(x, y) \quad (3.11)$$



Chapter 4 Results

4.1 The longitudinal model

4.1.1 Evaluation index

The training accuracy of longitudinal model is 99.5%, and its validation accuracy is 96.88% when the model was saved at the epoch 257. As mentioned in 3.10.1, validation accuracy would be the representative accuracy of this model. At the same epoch, the training loss of this model is 0.0135, and the validation loss is 0.13476. The accuracy and loss trends are shown in Figure 4-1.

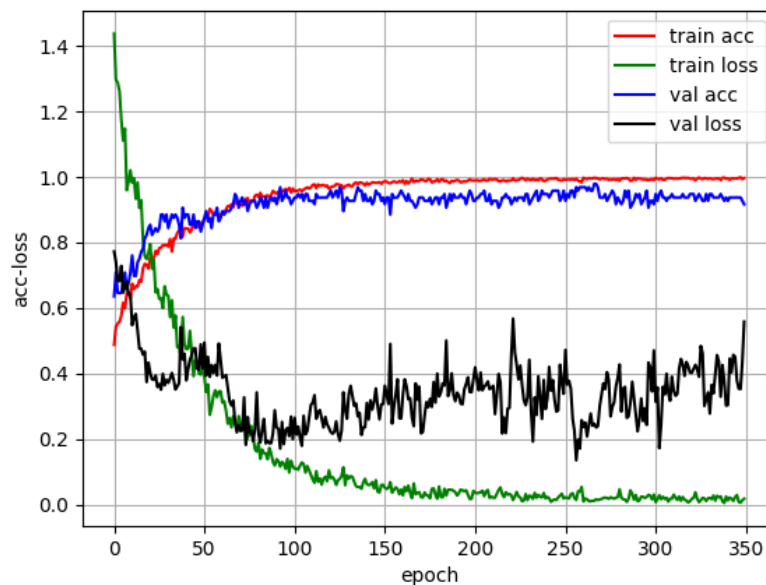


Figure 4-1: Learning curve of the longitudinal model.

4.1.1.1 Testing dataset with longitudinal view of supraspinatus tendon

Table 4-1 is the confusion matrix of longitudinal model testing against longitudinal view of supraspinatus tendon. True positive(TP) of this model is 116, and false positive(FP) is 6. False negative(FN) of this model is 14, and true negative(TN) is 124.

Table 4-1: Confusion matrix of the longitudinal model testing against longitudinal view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	116	6
	No calcification cases of supraspinatus tendons	14	124

Testing accuracy is calculated as 92.31% as shown in (4.1).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{116 + 124}{116 + 6 + 124 + 14} = 92.31\% \quad (4.1)$$

Sensitivity is calculated as 89.23% as shown in (4.2). Specificity is calculated as 95.38% as shown in (4.3).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{116}{116 + 14} = \frac{116}{130} = 89.23\% \quad (4.2)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{124}{6 + 124} = \frac{124}{130} = 95.38\% \quad (4.3)$$

PPV of this model is calculated as 95.08%, and NPV is 89.86% as shown in (4.4) and (4.5).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{116}{116 + 6} = \frac{116}{122} = 95.08\% \quad (4.4)$$

$$NPV = \frac{TN}{TN + FN} = \frac{124}{124 + 14} = \frac{124}{138} = 89.86\% \quad (4.5)$$

FPR is calculated as 4.62%, and FNR is 10.77% as shown in (4.6) and (4.7).

$$FPR = \frac{FP}{FP + TN} = 1 - Specificity = \frac{6}{6 + 124} = \frac{6}{130} = 4.62\% \quad (4.6)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{14}{116 + 14} = \frac{14}{130} = 10.77\% \quad (4.7)$$

+LR of this model is calculated as 19.33 in (4.8), which is a strong evidence to rule in disease in the clinical significance, and it means that the longitudinal model is useful in judging with the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon. -LR is 0.11 as shown in (4.9).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{116}{116 + 14}}{\frac{6}{6 + 124}} = \frac{\frac{116}{130}}{\frac{6}{130}} = \frac{116}{6} = 19.33 \quad (4.8)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{14}{116 + 14}}{\frac{124}{6 + 124}} = \frac{\frac{14}{130}}{\frac{124}{130}} = \frac{14}{124} = 0.11 \quad (4.9)$$

ROC curve of the longitudinal model testing against longitudinal view of supraspinatus tendon is shown in Figure 4-2. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against longitudinal view of supraspinatus tendon is calculated as 0.95, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

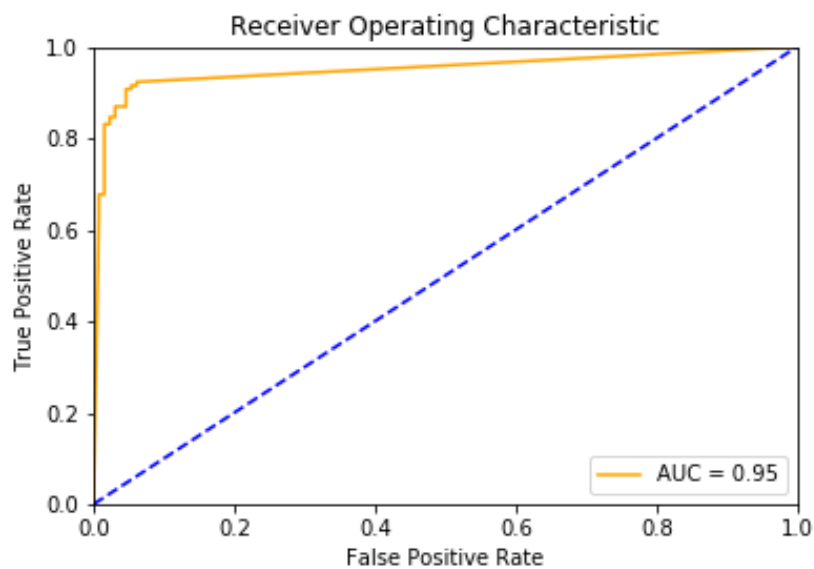


Figure 4-2: ROC curve of the longitudinal model testing against longitudinal view of supraspinatus tendon.

4.1.1.2 Testing dataset with transverse view of supraspinatus tendon

Table 4-2 is the confusion matrix of longitudinal model testing against transverse view of supraspinatus tendon. True positive(TP) of this model is 52, and false positive(FP) is 18. False negative(FN) of this model is 8, and true negative(TN) is 42.



Table 4-2: Confusion matrix of the longitudinal model testing against transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	52	18
	No calcification cases of supraspinatus tendons	8	42

Testing accuracy is calculated as 78.33% as shown in (4.10).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{52 + 42}{52 + 18 + 42 + 8} = 78.33\% \quad (4.10)$$

Sensitivity is calculated as 86.67% as shown in (4.11). Specificity is calculated as 70% as shown in (4.12).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{52}{52 + 8} = \frac{52}{60} = 86.67\% \quad (4.11)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{42}{18 + 42} = \frac{42}{60} = 70\% \quad (4.12)$$

PPV of this model is calculated as 74.29%, and NPV is 84% as shown in (4.13) and (4.14).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{52}{52 + 18} = \frac{52}{70} = 74.29\% \quad (4.13)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{42}{42 + 8} = \frac{42}{50} = 84\% \quad (4.14)$$

FPR is calculated as 30%, and FNR is 13.33% as shown in (4.15) and (4.16).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{18}{18 + 42} = \frac{18}{60} = 30\% \quad (4.15)$$

$$\text{FNR} = \frac{FN}{TP + FN} = 1 - \text{Sensitivity} = \frac{8}{52 + 8} = \frac{8}{60} = 13.33\% \quad (4.16)$$

+LR of this model is calculated as 2.89 in (4.17), which is a weak evidence to rule in disease in the clinical significance, and it means that the longitudinal model is able to judge with the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon, but its accuracy would not be precise enough. -LR is 0.19 as shown in (4.18).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{52}{52 + 8}}{\frac{18}{18 + 42}} = \frac{\frac{52}{60}}{\frac{24}{60}} = \frac{52}{24} = 2.89 \quad (4.17)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{8}{52 + 8}}{\frac{42}{18 + 42}} = \frac{\frac{8}{60}}{\frac{42}{60}} = \frac{8}{42} = 0.19 \quad (4.18)$$

ROC curve of the longitudinal model testing against transverse view of supraspinatus tendon is shown in Figure 4-3. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against transverse view of supraspinatus tendon is calculated as 0.89, this model is evaluated as an excellent discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

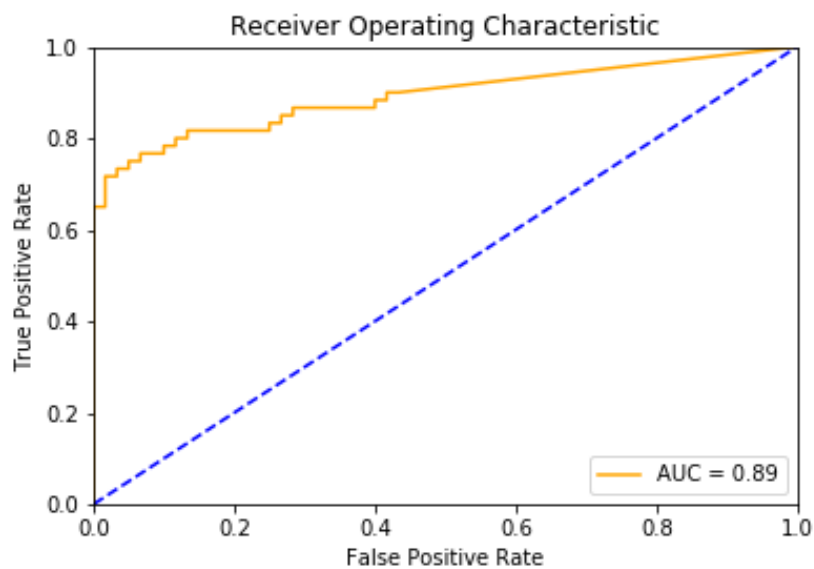


Figure 4-3: ROC curve of the longitudinal model testing against transverse view of supraspinatus tendon.

4.1.1.3 Testing dataset with both longitudinal and transverse view of supraspinatus tendon

Table 4-3 is the confusion matrix of longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon. True positive(TP) of this model is 168, and false positive(FP) is 24. False negative(FN) of this model is 22, and true negative(TN) is 166.

Table 4-3: Confusion matrix of the longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	168	24
	No calcification cases of supraspinatus tendons	22	166

Testing accuracy is calculated as 87.89% as shown in (4.19).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{168 + 166}{168 + 24 + 166 + 22} = 87.89\% \quad (4.19)$$

Sensitivity is calculated as 88.42% as shown in (4.20). Specificity is calculated as 87.37% as shown in (4.21).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{168}{168 + 22} = \frac{168}{190} = 88.42\% \quad (4.20)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{166}{24 + 166} = \frac{166}{190} = 87.37\% \quad (4.21)$$

PPV of this model is calculated as 87.5%, and NPV is 88.3% as shown in (4.22) and (4.23).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{168}{168 + 24} = \frac{168}{192} = 87.5\% \quad (4.22)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{166}{166 + 22} = \frac{166}{188} = 88.3\% \quad (4.23)$$

FPR is calculated as 12.63%, and FNR is 11.58% as shown in (4.24) and (4.25).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{24}{24 + 166} = \frac{24}{190} = 12.63\% \quad (4.24)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{22}{168 + 22} = \frac{22}{190} = 11.58\% \quad (4.25)$$

+LR of this model is calculated as 7.0 in (4.26), which is a moderate evidence to rule in disease in the clinical significance, and it means that the longitudinal model is useful in judging with the existence of supraspinatus calcific tendinopathy of longitudinal view and transverse view of supraspinatus tendon. -LR is 0.13 as shown in (4.27).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{168}{168 + 22}}{\frac{24}{24 + 166}} = \frac{\frac{168}{190}}{\frac{24}{190}} = \frac{168}{24} = 7.0 \quad (4.26)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{22}{168 + 22}}{\frac{166}{24 + 166}} = \frac{\frac{22}{190}}{\frac{166}{190}} = \frac{22}{166} = 0.13 \quad (4.27)$$

ROC curve of the longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon is shown in Figure 4-4. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against longitudinal view and transverse view of supraspinatus tendon is calculated as 0.93, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view and transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

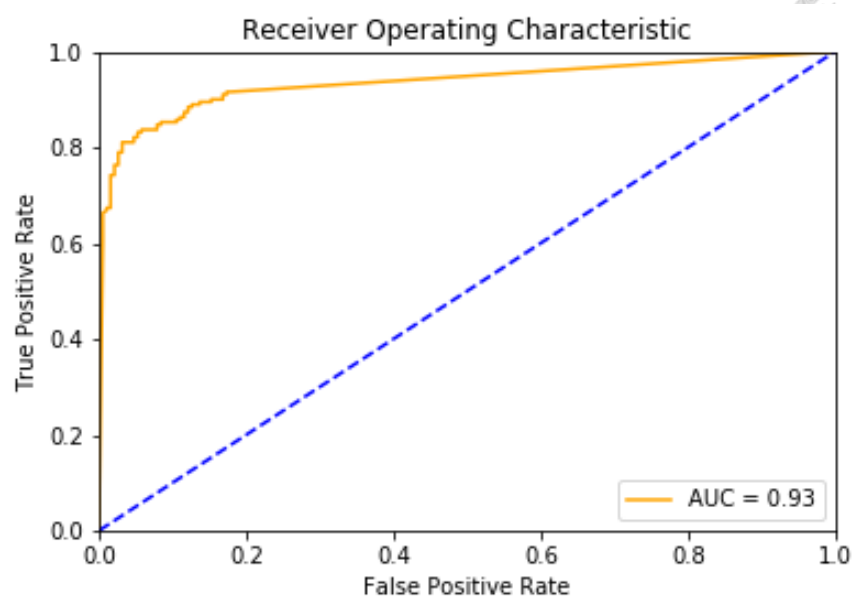


Figure 4-4: ROC curve of the longitudinal model testing against longitudinal view and transverse view of supraspinatus tendon.

4.1.2 Heatmap

4.1.2.1 Testing dataset with longitudinal view of supraspinatus tendon

Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon are shown in Figure 4-5. They visualized how the longitudinal model judge the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon.

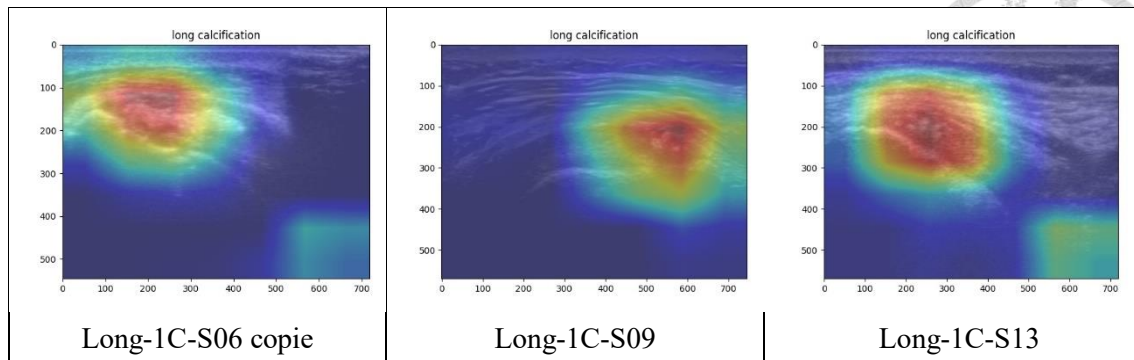


Figure 4-5: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the longitudinal model.

4.1.2.2 Testing dataset with transverse view of supraspinatus tendon

Representative heatmaps of testing dataset with transverse view of supraspinatus tendon are shown in Figure 4-6. They visualized how the longitudinal model judge the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon.

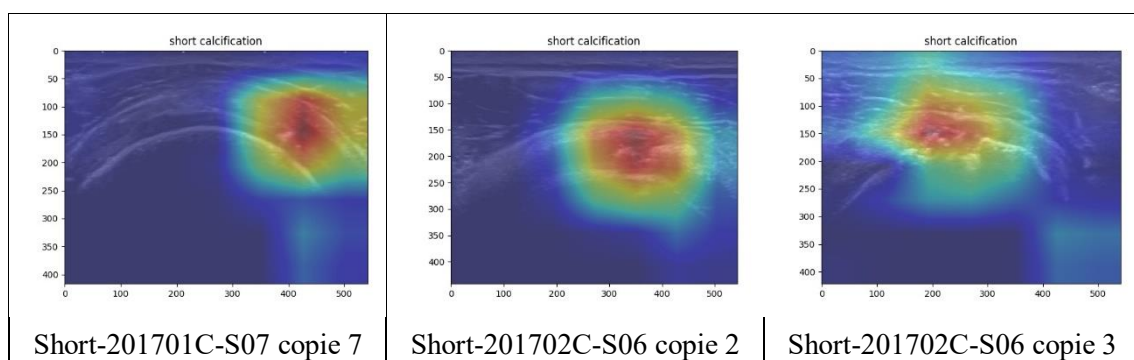


Figure 4-6: Representative heatmaps of testing dataset with transverse view of supraspinatus tendon in the longitudinal model.



4.2 The transverse model

4.2.1 Evaluation index

The training accuracy of the transverse model is 99.86%, and its validation accuracy is 94.64% when the model was saved at the epoch 339. At the same epoch, the training loss of this model is 0.01, and the validation loss is 0.2028. The accuracy and loss trends are shown in Figure 4-7.

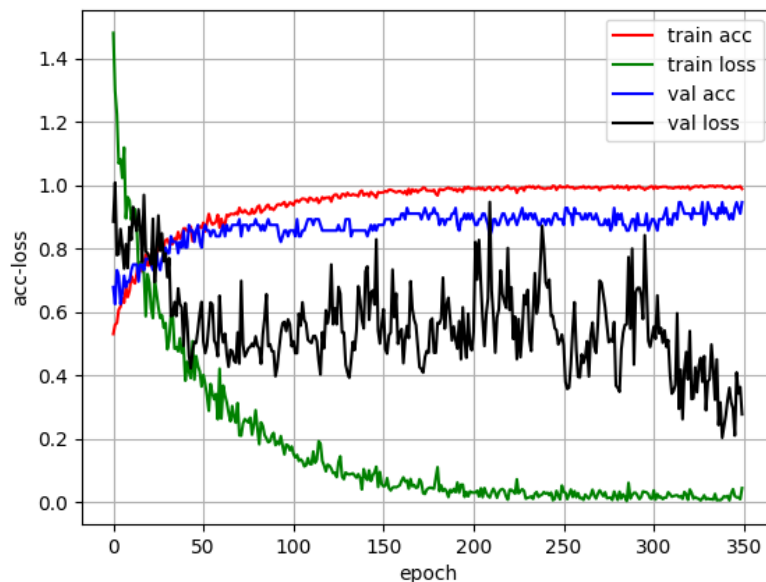


Figure 4-7: Learning curve of the transverse model.

4.2.1.1 Testing dataset with longitudinal view of supraspinatus tendon

Table 4-4 is the confusion matrix of the transverse model testing against longitudinal view of supraspinatus tendon. True positive(TP) of this model is 115, and false positive(FP) is 65. False negative(FN) of this model is 15, and true negative(TN) is 65.

Table 4-4: Confusion matrix of the transverse model testing against longitudinal view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	115	65
	No calcification cases of supraspinatus tendons	15	65

Testing accuracy is calculated as 69.23% as shown in (4.28).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{115 + 65}{115 + 65 + 65 + 15} = 69.23\% \quad (4.28)$$

Sensitivity is calculated as 88.46% as shown in (4.29). Specificity is calculated as 50% as shown in (4.30).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{115}{115 + 15} = \frac{115}{130} = 88.46\% \quad (4.29)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{65}{65 + 65} = \frac{65}{130} = 50\% \quad (4.30)$$

PPV of this model is calculated as 63.89%, and NPV is 81.25% as shown in (4.31) and (4.32).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{115}{115 + 65} = \frac{115}{180} = 63.89\% \quad (4.31)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{65}{65 + 15} = \frac{65}{80} = 81.25\% \quad (4.32)$$

FPR is calculated as 50%, and FNR is 11.54% as shown in (4.33) and (4.34).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{65}{65 + 65} = \frac{65}{130} = 50\% \quad (4.33)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{15}{115 + 15} = \frac{15}{130} = 11.54\% \quad (4.34)$$

+LR of this model is calculated as 1.77 in (4.35), which has no significant change in the likelihood to rule in disease in the clinical significance, and it means that the transverse model is useless to judge the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon. -LR is 0.23 as shown in (4.36).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{115}{115 + 15}}{\frac{65}{65 + 65}} = \frac{\frac{115}{130}}{\frac{65}{130}} = 1.77 \quad (4.35)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{15}{115 + 15}}{\frac{65}{65 + 65}} = \frac{\frac{15}{130}}{\frac{65}{130}} = 0.23 \quad (4.36)$$

ROC curve of the transverse model testing against longitudinal view of supraspinatus tendon is shown in Figure 4-8. The ROC curve is located in the left region which stands for a good classifier. Since AUC of this model testing against longitudinal view of supraspinatus tendon is calculated as 0.8, this model is evaluated as an excellent discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

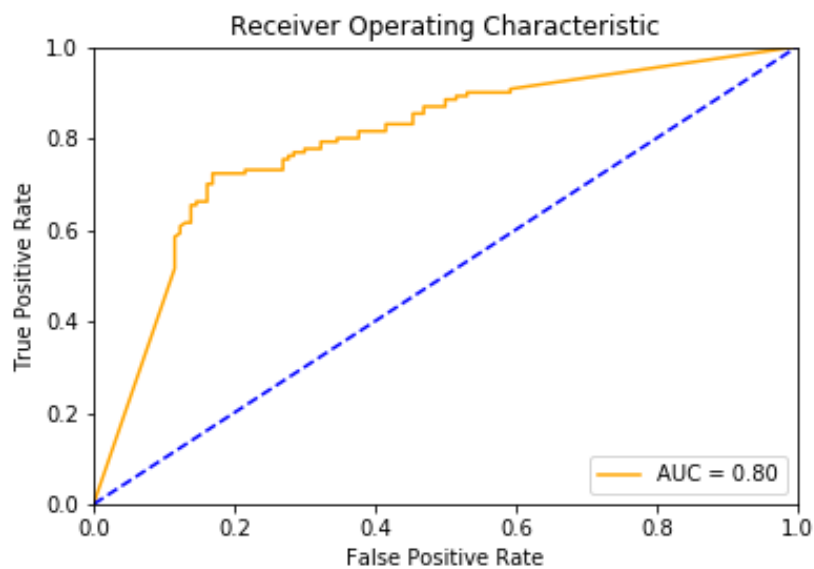


Figure 4-8: ROC curve of the transverse model testing against longitudinal view of supraspinatus tendon.

4.2.1.2 Testing dataset with transverse view of supraspinatus tendon

Table 4-5 is the confusion matrix of the transverse model testing against transverse view of supraspinatus tendon. True positive(TP) of this model is 50, and false positive(FP) is 3. False negative(FN) of this model is 10, and true negative(TN) is 57.

Table 4-5: Confusion matrix of the transverse model testing against transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	50	3
	No calcification cases of supraspinatus tendons	10	57

Testing accuracy is calculated as 89.17% as shown in (4.37).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{50 + 57}{50 + 3 + 57 + 10} = 89.17\% \quad (4.37)$$

Sensitivity is calculated as 83.33% as shown in (4.38). Specificity is calculated as 95% as shown in (4.39).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{50}{50 + 10} = \frac{50}{60} = 83.33\% \quad (4.38)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{57}{3 + 57} = \frac{57}{60} = 95\% \quad (4.39)$$

PPV of this model is calculated as 94.34%, and NPV is 85.07% as shown in (4.40) and (4.41).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{50}{50 + 3} = \frac{50}{53} = 94.34\% \quad (4.40)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{57}{57 + 10} = \frac{57}{67} = 85.07\% \quad (4.41)$$

FPR is calculated as 5%, and FNR is 16.67% as shown in (4.42) and (4.43).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{3}{65 + 57} = \frac{3}{122} = 2.46\% \quad (4.42)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{10}{50 + 10} = \frac{10}{60} = 16.67\% \quad (4.43)$$

+LR of this model is calculated as 16.67 in (4.44), which is a strong evidence to rule in disease in the clinical significance, and it means that the transverse model is able to judge with the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon. -LR is 0.18 as shown in (4.45).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{50}{50 + 10}}{\frac{3}{3 + 57}} = \frac{\frac{50}{60}}{\frac{3}{60}} = 16.67 \quad (4.44)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{10}{50 + 10}}{\frac{57}{3 + 57}} = \frac{\frac{10}{60}}{\frac{57}{60}} = 0.18 \quad (4.45)$$

ROC curve of the transverse model testing against transverse view of supraspinatus tendon is shown in Figure 4-9. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against transverse view of supraspinatus tendon is calculated as 0.93, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

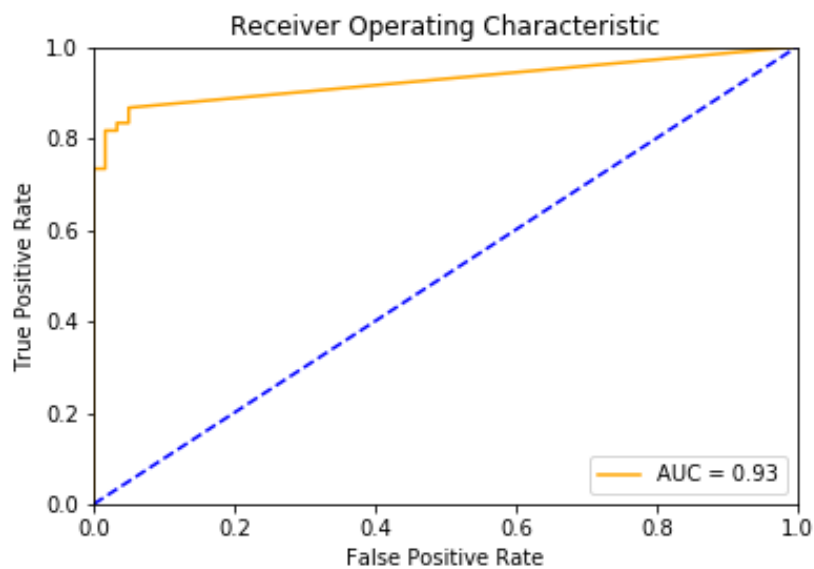


Figure 4-9: ROC curve of the transverse model testing against transverse view of supraspinatus tendon.

4.2.1.3 Testing dataset with both longitudinal and transverse view of supraspinatus tendon

Table 4-6 is the confusion matrix of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon. True positive(TP) of this model is 165, and false positive(FP) is 68. False negative(FN) of this model is 25, and true negative(TN) is 122.

Table 4-6: Confusion matrix of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	165	68
	No calcification cases of supraspinatus tendons	25	122

Testing accuracy is calculated as 75.53% as shown in (4.46).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{165 + 122}{165 + 68 + 122 + 25} = 75.53\% \quad (4.46)$$

Sensitivity is calculated as 86.84% as shown in (4.47). Specificity is calculated as 64.21% as shown in (4.48).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{165}{165 + 25} = \frac{165}{190} = 86.84\% \quad (4.47)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{122}{68 + 122} = \frac{122}{190} = 64.21\% \quad (4.48)$$

PPV of this model is calculated as 70.82%, and NPV is 82.99% as shown in (4.49) and (4.50).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{165}{165 + 68} = \frac{165}{233} = 70.82\% \quad (4.49)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{122}{122 + 25} = \frac{122}{147} = 82.99\% \quad (4.50)$$

FPR is calculated as 35.79%, and FNR is 13.16% as shown in (4.51) and (4.52).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{68}{68 + 122} = \frac{68}{190} = 35.79\% \quad (4.51)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{25}{165 + 25} = \frac{25}{180} = 13.16\% \quad (4.52)$$

+LR of this model is calculated as 2.43 in (4.53), which is a weak evidence to rule in disease in the clinical significance, and it means that the transverse model is able to judge with the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon, but its accuracy would not be precise enough. -LR is 0.2 as shown in (4.54).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{165}{165 + 25}}{\frac{68}{68 + 122}} = \frac{\frac{165}{190}}{\frac{68}{190}} = 2.43 \quad (4.53)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{25}{165 + 25}}{\frac{122}{68 + 122}} = \frac{\frac{25}{190}}{\frac{122}{190}} = 0.2 \quad (4.54)$$

ROC curve of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon is shown in Figure 4-10. The ROC curve is located in the left region which stands for a good classifier. Since AUC of this model testing against longitudinal view and transverse view of supraspinatus tendon is calculated as 0.84, this model is evaluated as an excellent discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view and transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

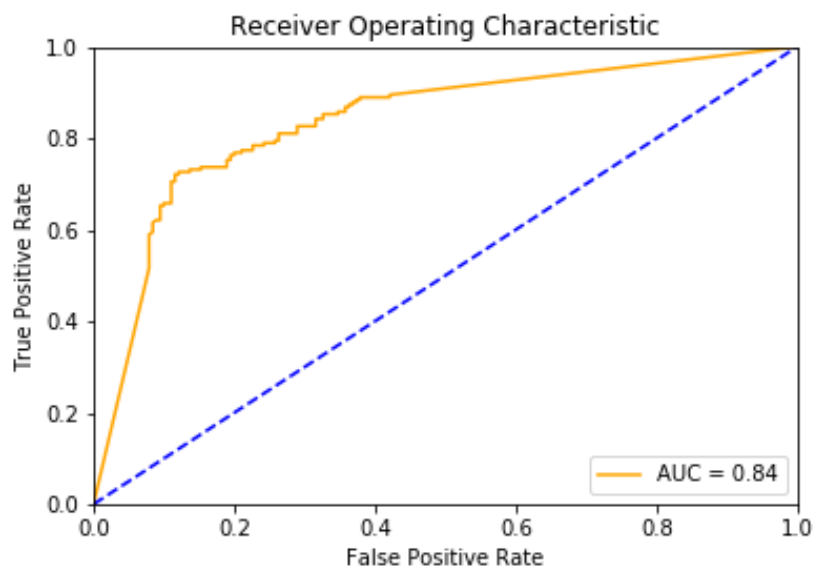


Figure 4-10: ROC curve of the transverse model testing against longitudinal view and transverse view of supraspinatus tendon.

4.2.2 Heatmap

4.2.2.1 Testing dataset with longitudinal view of supraspinatus tendon

Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon are shown in Figure 4-11. They visualized how the transverse model judge the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon.

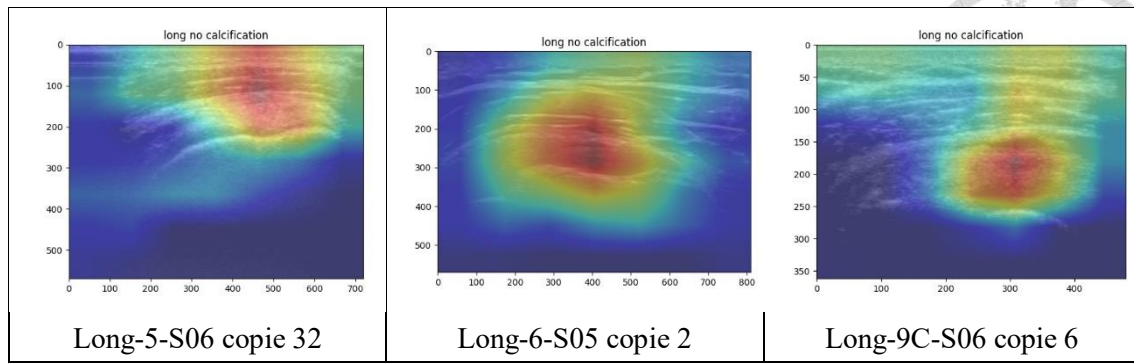


Figure 4-11: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the transverse model.

4.2.2.2 Testing dataset with transverse view of supraspinatus tendon

Representative heatmaps of testing dataset with transverse view of supraspinatus tendon are shown in Figure 4-12. They visualized how the transverse model judge the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon.

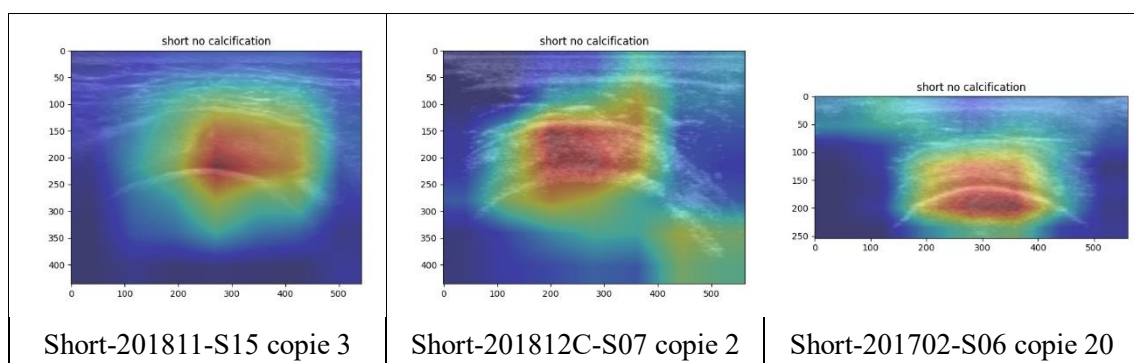


Figure 4-12: Representative heatmaps of testing dataset with transverse view of supraspinatus tendon in the transverse model.



4.3 The longi-trans model

4.3.1 Evaluation index

The training accuracy of the longi-trans model is 96.72%, and its validation accuracy is 90.62% when the model was saved at the epoch 313. At the same epoch, the training loss of this model is 0.034, and the validation loss is 0.4194. The accuracy and loss trends are shown in Figure 4-13.

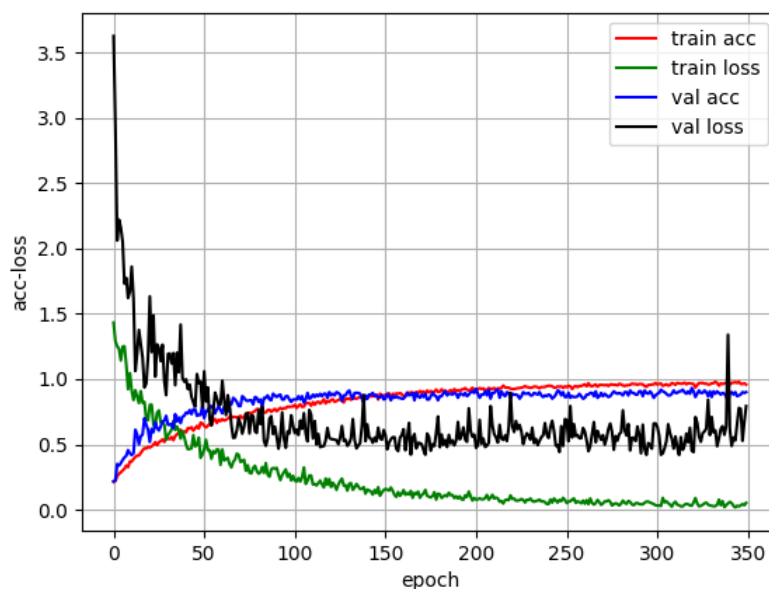


Figure 4-13: Learning curve of the longi-trans model.

4.3.1.1 Testing dataset with longitudinal view of supraspinatus tendon

Table 4-7 is the confusion matrix of the longi-trans model testing against longitudinal view of supraspinatus tendon. True positive(TP) of this model is 115, and



false positive(FP) is 7. False negative(FN) of this model is 15, and true negative(TN) is 123.

Table 4-7: Confusion matrix of the longi-trans model testing against longitudinal view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	115	7
	No calcification cases of supraspinatus tendons	15	123

Testing accuracy is calculated as 91.54% as shown in (4.55).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{115 + 123}{115 + 7 + 123 + 15} = 91.54\% \quad (4.55)$$

Sensitivity is calculated as 88.46% as shown in (4.56). Specificity is calculated as 94.62% as shown in (4.57).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{115}{115 + 15} = \frac{115}{130} = 88.46\% \quad (4.56)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{123}{7 + 123} = \frac{123}{130} = 94.62\% \quad (4.57)$$

PPV of this model is calculated as 94.26%, and NPV is 89.13% as shown in (4.58) and (4.59).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{115}{115 + 7} = \frac{115}{122} = 94.26\% \quad (4.58)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{123}{123 + 15} = \frac{123}{138} = 89.13\% \quad (4.59)$$

FPR is calculated as 5.38%, and FNR is 11.54% as shown in (4.60) and (4.61).

$$FPR = \frac{FP}{FP + TN} = 1 - Specificity = \frac{7}{7 + 123} = \frac{7}{129} = 5.38\% \quad (4.60)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{15}{115 + 15} = \frac{15}{130} = 11.54\% \quad (4.61)$$

+LR of this model is calculated as 16.43 in (4.62), which is a strong evidence to rule in disease in the clinical significance, and it means that the longi-trans model is a good at judging with the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon. -LR is 0.12 as shown in (4.63).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{115}{115 + 15}}{\frac{7}{7 + 123}} = \frac{\frac{115}{130}}{\frac{7}{130}} = \frac{115}{7} = 16.43 \quad (4.62)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{15}{115 + 15}}{\frac{123}{7 + 123}} = \frac{\frac{15}{130}}{\frac{123}{130}} = \frac{15}{123} = 0.12 \quad (4.63)$$

ROC curve of the longi-trans model is shown in Figure 4-14. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against longitudinal view of supraspinatus tendon is calculated as 0.95, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

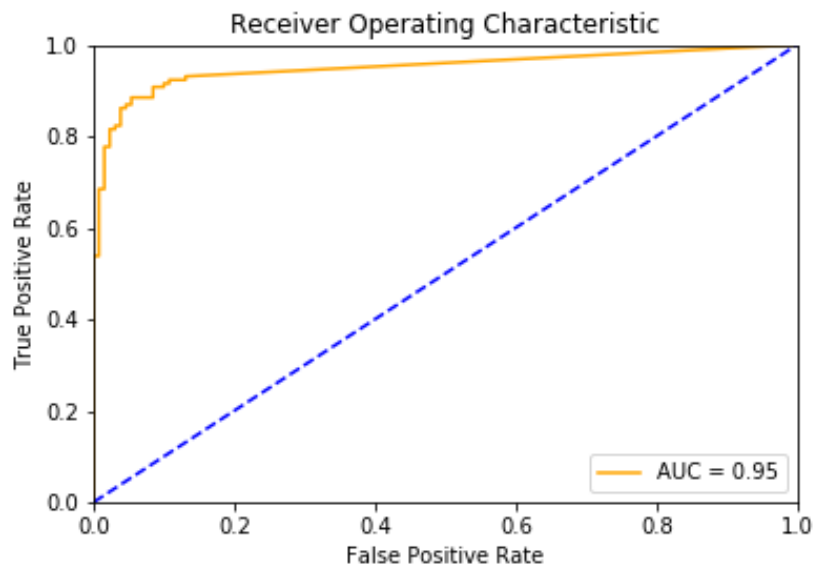


Figure 4-14: ROC curve of the longi-trans model testing against longitudinal view of supraspinatus tendon.

4.3.1.2 Testing dataset with transverse view of supraspinatus tendon

Table 4-8 is the confusion matrix of the longi-trans model testing against transverse view of supraspinatus tendon. True positive(TP) of this model is 52, and false positive(FP) is 3. False negative(FN) of this model is 8, and true negative(TN) is 57.

Table 4-8: Confusion matrix of the longi-trans model testing against transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	52	3
	No calcification cases of supraspinatus tendons	8	57

Testing accuracy is calculated as 90.83% as shown in (4.64).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{52 + 57}{52 + 3 + 57 + 8} = 90.83\% \quad (4.64)$$

Sensitivity is calculated as 86.67% as shown in (4.65). Specificity is calculated as 95% as shown in (4.66).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{52}{52 + 8} = \frac{52}{60} = 86.67\% \quad (4.65)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{57}{3 + 57} = \frac{57}{60} = 95\% \quad (4.66)$$

PPV of this model is calculated as 94.55%, and NPV is 87.69% as shown in (4.67) and (4.68).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{52}{52 + 3} = \frac{52}{55} = 94.55\% \quad (4.67)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{57}{57 + 8} = \frac{57}{65} = 87.69\% \quad (4.68)$$

FPR is calculated as 5%, and FNR is 13.33% as shown in (4.69) and (4.70).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{3}{3 + 57} = \frac{3}{60} = 5\% \quad (4.69)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{8}{52 + 8} = \frac{8}{60} = 13.33\% \quad (4.70)$$

+LR of this model is calculated as 17.33 in (4.71), which is a strong evidence to rule in disease in the clinical significance, and it means that the longi-trans model is a good at judging with the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon. -LR is 0.14 as shown in (4.72).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{52}{52 + 8}}{\frac{3}{3 + 57}} = \frac{\frac{52}{60}}{\frac{3}{60}} = \frac{52}{3} = 17.33 \quad (4.71)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{8}{52 + 8}}{\frac{57}{3 + 57}} = \frac{\frac{8}{60}}{\frac{57}{60}} = \frac{8}{57} = 0.14 \quad (4.72)$$

ROC curve of the longi-trans model testing against transverse view of supraspinatus tendon is shown in Figure 4-15. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against transverse view of supraspinatus tendon is calculated as 0.94, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

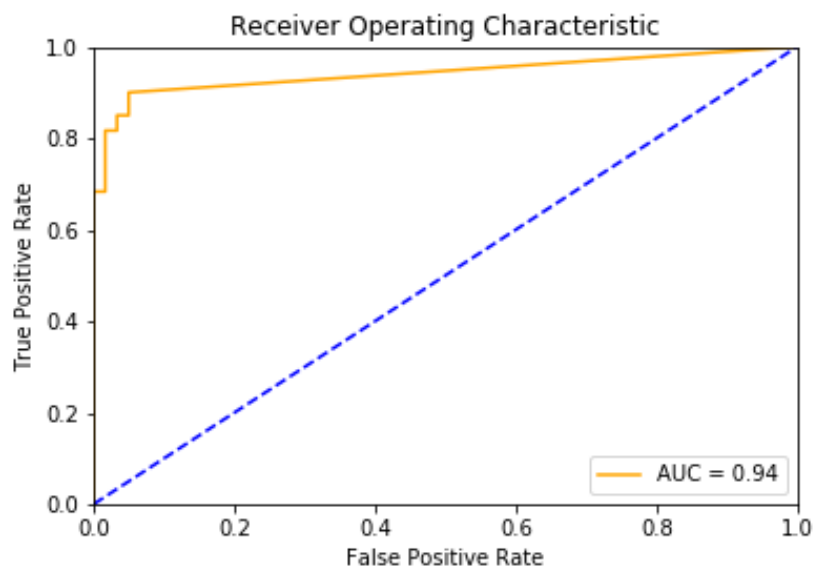


Figure 4-15: ROC curve of the longi-trans model testing against transverse view of supraspinatus tendon.

4.3.1.3 Testing dataset with longi-trans view of supraspinatus tendon

Table 4-9 is the confusion matrix of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon. True positive(TP) of this model is 167, and false positive(FP) is 10. False negative(FN) of this model is 23, and true negative(TN) is 180.



Table 4-9: Confusion matrix of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon.

		Actual classes	
		Calcification cases of supraspinatus tendons	No calcification cases of supraspinatus tendons
Predicted classes	Calcification cases of supraspinatus tendons	167	10
	No calcification cases of supraspinatus tendons	23	180

Testing accuracy is calculated as 91.32% as shown in (4.73).

$$\text{Testing accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{167 + 180}{167 + 10 + 180 + 23} = 91.32\% \quad (4.73)$$

Sensitivity is calculated as 87.89% as shown in (4.74). Specificity is calculated as 94.74% as shown in (4.75).

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{167}{167 + 23} = \frac{167}{190} = 87.89\% \quad (4.74)$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{180}{10 + 180} = \frac{180}{190} = 94.74\% \quad (4.75)$$

PPV of this model is calculated as 94.35%, and NPV is 88.67% as shown in (4.76) and (4.77).

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{167}{167 + 10} = \frac{167}{177} = 94.35\% \quad (4.76)$$

$$\text{NPV} = \frac{TN}{TN + FN} = \frac{180}{180 + 23} = \frac{180}{203} = 88.67\% \quad (4.77)$$

FPR is calculated as 5.26%, and FNR is 12.11% as shown in (4.78) and (4.79).

$$\text{FPR} = \frac{FP}{FP + TN} = 1 - \text{Specificity} = \frac{10}{10 + 180} = \frac{10}{190} = 5.26\% \quad (4.78)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Sensitivity = \frac{23}{167 + 23} = \frac{23}{190} = 12.11\% \quad (4.79)$$

+LR of this model is calculated as 16.7 in (4.80), which is a strong evidence to rule in disease in the clinical significance, and it means that the longi-trans model is a good at judging with the existence of supraspinatus calcific tendinopathy of longitudinal view and transverse view of supraspinatus tendon. -LR is 0.13 as shown in (4.81).

$$+LR = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{167}{167 + 23}}{\frac{10}{10 + 180}} = \frac{\frac{167}{190}}{\frac{10}{190}} = \frac{167}{10} = 16.7 \quad (4.80)$$

$$-LR = \frac{\frac{FN}{TP + FN}}{\frac{TN}{FP + TN}} = \frac{\frac{23}{167 + 23}}{\frac{180}{10 + 180}} = \frac{\frac{23}{190}}{\frac{180}{190}} = \frac{23}{180} = 0.13 \quad (4.81)$$

ROC curve of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon is shown in Figure 4-16. The ROC curve is located in the upper left region which stands for a good classifier. Since AUC of this model testing against longitudinal view and transverse view of supraspinatus tendon is calculated as 0.95, this model is evaluated as an outstanding discrimination for the judgement of the existence of supraspinatus calcific tendinopathy of longitudinal view and transverse view of supraspinatus tendon by the meaning of AUC value mentioned in 3.10.14.

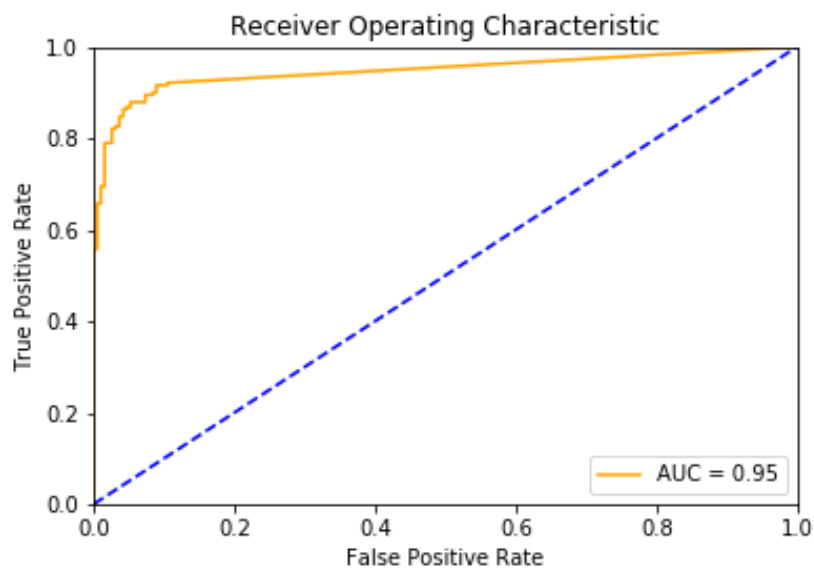


Figure 4-16: ROC curve of the longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon.

4.3.2 Heatmap

4.3.2.1 Testing dataset with longitudinal view of supraspinatus tendon

Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon are shown in Figure 4-17. They visualized how the longi-trans model judge the existence of supraspinatus calcific tendinopathy of longitudinal view of supraspinatus tendon.

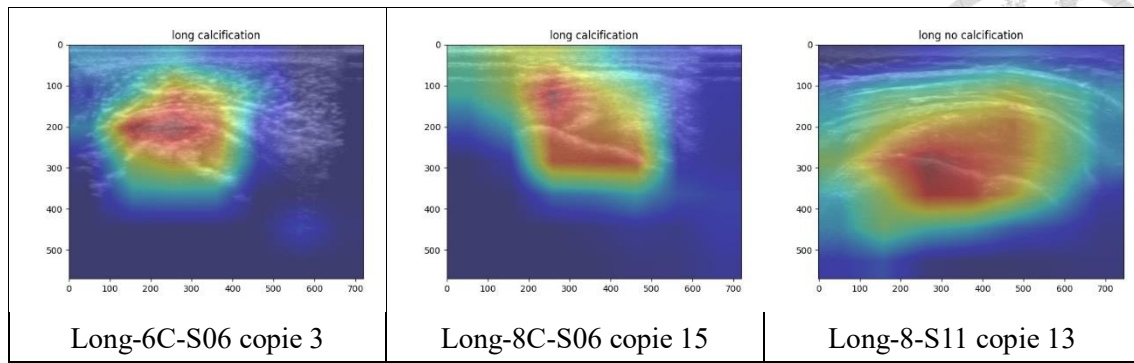


Figure 4-17: Representative heatmaps of testing dataset with longitudinal view of supraspinatus tendon in the longi-trans model.

4.3.2.2 Testing dataset with transverse view of supraspinatus tendon

Representative heatmaps of testing dataset with transverse view of supraspinatus tendon are shown in Figure 4-18. They visualized how the longi-trans model judge the existence of supraspinatus calcific tendinopathy of transverse view of supraspinatus tendon.

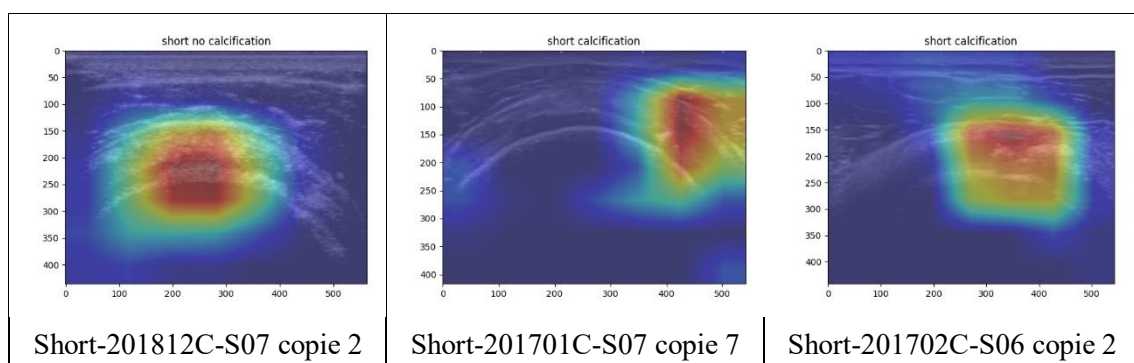


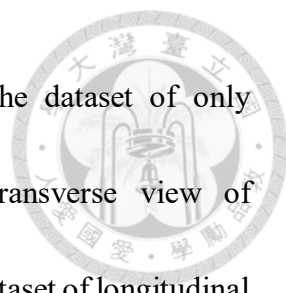
Figure 4-18: Representative heatmaps of testing dataset with transverse view of supraspinatus tendon in the longi-trans model.

Chapter 5 Discussion



5.1 Evaluation index

Figure 5-1 is the testing accuracy between three models against three different distribution of testing dataset. For the longitudinal model, its testing accuracy against the longitudinal view of supraspinatus tendon was much better than the transverse view of supraspinatus tendon since the longitudinal model was trained on data of longitudinal view of supraspinatus tendon. However, there were still 78.33% of the transverse view of supraspinatus tendon in the testing dataset correctly diagnosed by the longitudinal model. It shows that the longitudinal model diagnoses the existence of supraspinatus calcific tendinopathy in a certain level of ability. It learned some clues form the longitudinal view of supraspinatus tendon which also appears in the transverse view of supraspinatus tendon. For the transverse model, its testing accuracy against the transverse view of supraspinatus tendon is 89.17% which is 19.94% higher than the testing accuracy against the longitudinal view of supraspinatus tendon. Although the performance of the transverse model is not as good as the longitudinal model, it is still found that a model which learned only one view of supraspinatus tendon, is still able to diagnose the existence of supraspinatus calcific tendinopathy to a certain extent no matter the view of supraspinatus tendon. As for longi-trans model, attributed to the multiple views of supraspinatus tendon



it learned, this model has the highest testing accuracy against the dataset of only transverse view and the dataset with both longitudinal and transverse view of supraspinatus tendon, and a little lower testing accuracy against the dataset of longitudinal view of supraspinatus tendon. Testing accuracy among different models and testing dataset shows that a model based on only one view of supraspinatus tendon diagnoses the existence of supraspinatus calcific tendinopathy in a certain level of ability, but a model based on both longitudinal and transverse view of supraspinatus tendon has a better performance in reality since longitudinal view and transverse view of supraspinatus tendon have the same contribution to the judgement of the existence of supraspinatus calcific tendinopathy. It was also found that multiple view of supraspinatus tendon would not contribute to the confusion in models, but a good performance still based on the quantity and the richness of input data.

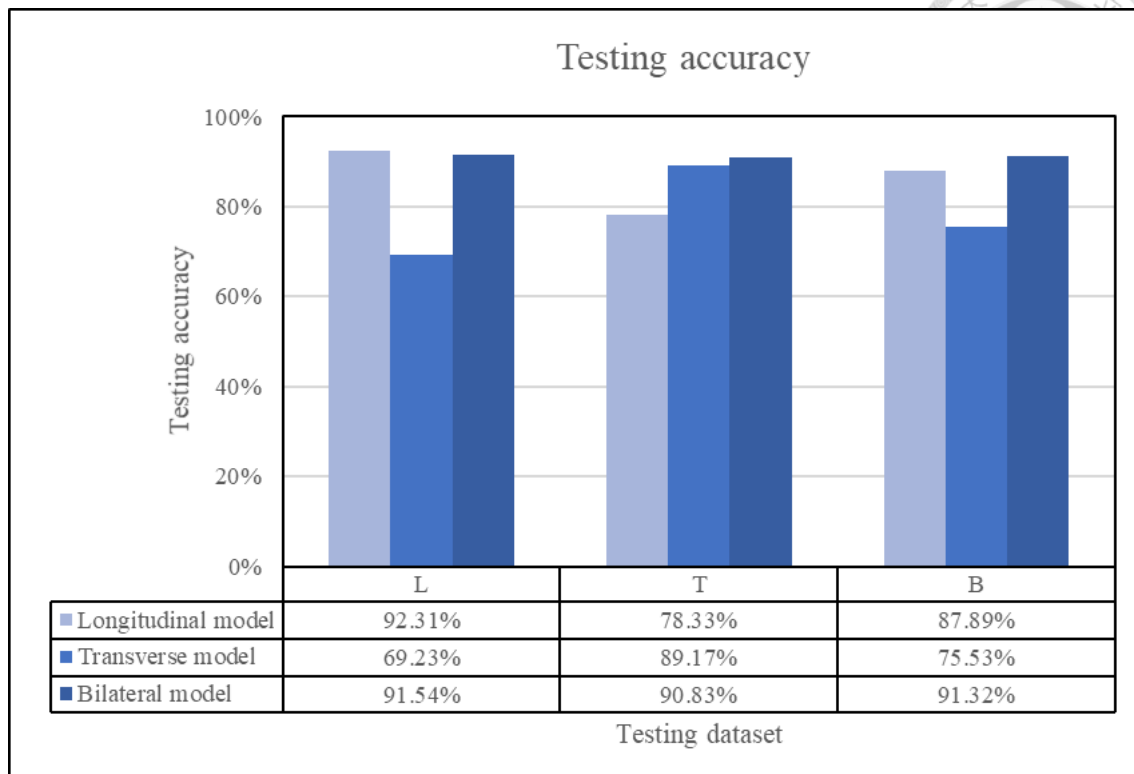


Figure 5-1: The testing accuracy between three models against three different distribution of testing dataset.

In addition to the testing accuracy against three different distribution of testing dataset, sensitivity and specificity also shows the different ability against these different distribution of testing dataset in the longitudinal model, the transverse model, and the longi-trans model. Figure 5-2 is the sensitivity between three models against three different distribution of testing dataset. It was found that there is no obvious difference among the different combination of models and testing dataset. However, their specificity shown in Figure 5-3 put forward a different point of view. Specificity of the model testing against a never seen view of supraspinatus tendon is much lower than a seen view of

supraspinatus tendon. It shows that the longi-trans model is needed for good performances of the existence of supraspinatus calcific tendinopathy against both the longitudinal view and transverse view of supraspinatus tendon.

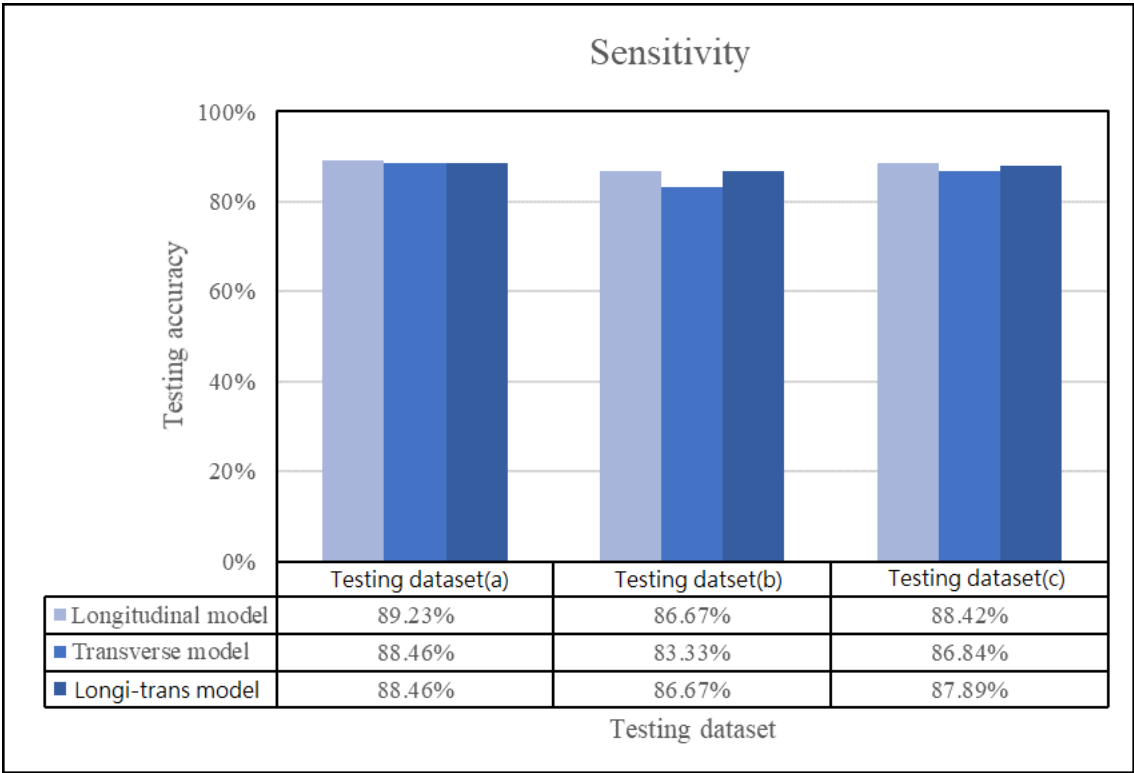


Figure 5-2: Sensitivity between three models against three different distribution of testing dataset.

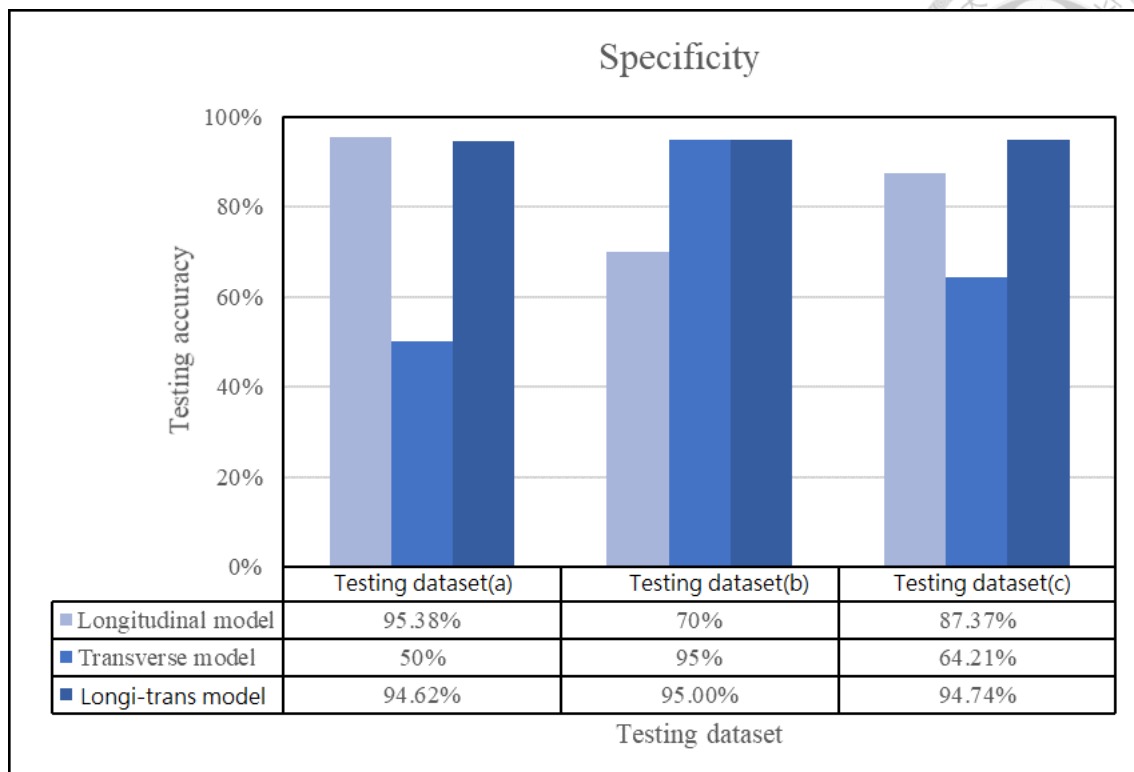
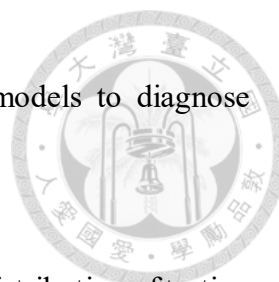


Figure 5-3: Specificity between three models against three different distribution of testing dataset.

Table 5-1 is the comparison between three models against three different distribution of testing dataset. A +LR over 10 happens in the longitudinal model against longitudinal view of supraspinatus tendon, the transverse model against transverse view of supraspinatus tendon, and the longi-trans model against both longitudinal view and transverse view of supraspinatus tendon. +LR of model tested against a never seen view of supraspinatus tendon is around 2.5 which is not useless to the diagnose, but weak. Higher AUC of the models testing against seen view of supraspinatus tendon also show



that the longi-trans model is the best model among these three models to diagnose longitudinal view and transverse view of supraspinatus tendon.

Table 5-1: Comparison between three models against three different distribution of testing

dataset. The first alphabet stands for the model category: “L” is the longitudinal model, “T” is the transverse model, “LT” is the longi-trans model. The second alphabet stands for the testing dataset category: “L” is the testing dataset with longitudinal view of supraspinatus tendon, “T” is the testing dataset with transverse view of supraspinatus tendon, “LT” is the testing dataset with both longitudinal and transverse view of supraspinatus tendon.

	L-L	L-T	L-LT	T-L	T-T	T-LT	LT-L	LT-T	LT-LT
Accuracy	99.5%			99.86%			96.72%		
Validation accuracy	96.88%			94.64%			90.62%		
Loss	0.0135			0.01			0.034		
Validation loss	0.13476			0.2028			0.4194		
Testing accuracy	92.31%	78.33%	87.89%	69.23%	89.17%	75.53%	91.54 %	90.83 %	91.32 %
Sensitivity	89.23%	86.67%	88.42%	88.46%	83.33%	86.84%	88.46 %	86.67 %	87.89 %
Specificity	95.38%	70%	87.37%	50%	95%	64.21%	94.62 %	95.0 %	94.74 %
PPV	95.08%	74.29%	87.5%	63.89%	94.34%	70.82%	94.26 %	94.55 %	94.35 %
NPV	89.86%	84%	88.3%	81.25%	85.07%	82.99%	89.13 %	87.69 %	88.67 %
FPR	4.62%	30%	12.63%	50%	5.0%	35.79%	5.38 %	5.0 %	5.26 %
FNR	10.77%	13.33%	11.58%	11.54%	16.67%	13.16%	11.54 %	13.33 %	12.11 %
+LR	19.33	2.89	7	1.77	16.67	2.43	16.43	17.33	16.7
-LR	0.11	0.19	0.13	0.23	0.18	0.2	0.12	0.14	0.13
AUC	0.95	0.89	0.93	0.84	0.93	0.84	0.95	0.94	0.95

5.2 Heatmap



The heatmaps of these three models showed that they learned some different features of the ultrasound images individually. The color distribution in each heatmap can be classified as 3 groups except those heatmaps with no notable color distribution:

- (1) On supraspinatus tendon: Heatmaps with color distributed on supraspinatus tendon are shown in Figure 5-4. In these heatmaps, color distributed on supraspinatus tendon which was a basis to inference that supraspinatus tendon would be contributing to the judgement against the existence of supraspinatus calcific tendinopathy in the models.

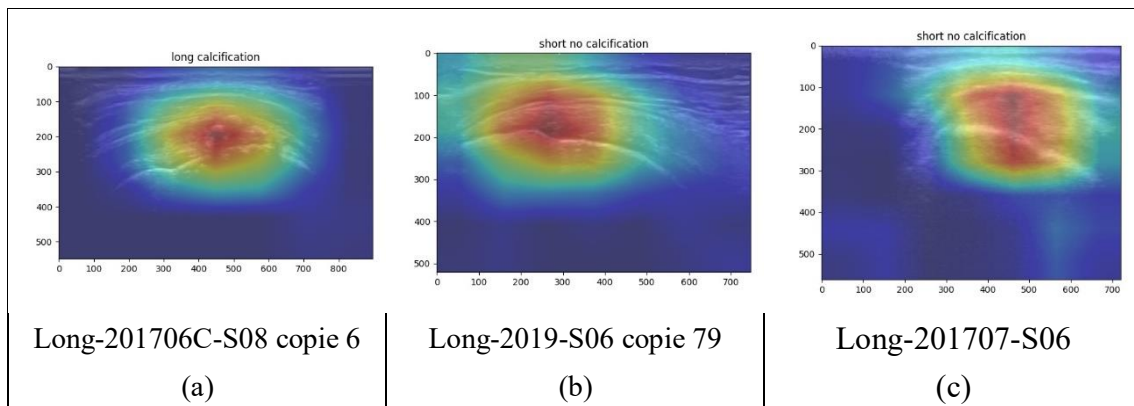


Figure 5-4: Heatmaps with color distributed on supraspinatus tendon. (a) A heatmap of the longitudinal model. (b) A heatmap of the transverse model. (c) A heatmap of the longi-trans model.

- (2) Under supraspinatus tendon: Heatmaps with color distributed under supraspinatus tendon are shown in Figure 5-5. In these heatmaps, color

distributed under supraspinatus tendon or closed to the bottom edge of it which was a basis to inference that the edge of supraspinatus tendon, especially the bottom of supraspinatus tendon, would be a contribution to the judgement against the existence of supraspinatus calcific tendinopathy in the models.

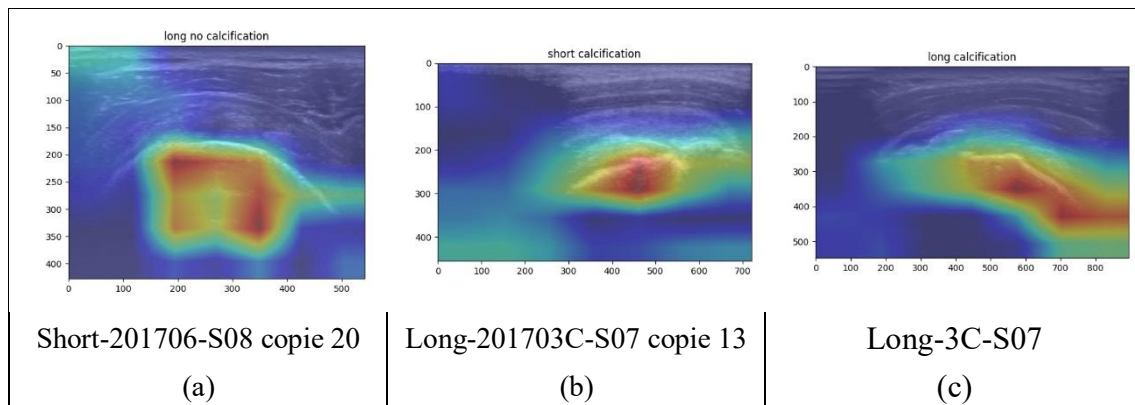


Figure 5-5: Heatmaps with color distributed under supraspinatus tendon. (a) A heatmap of the longitudinal model. (b) A heatmap of the transverse model. (c) A heatmap of the longi-trans model.

- (3) Greater tuberosity: Heatmaps with color distributed in greater tuberosity are shown in Figure 5-6. In these heatmaps, color distributed in greater tuberosity which was a basis to inference that greater tuberosity would also be a contribution to the judgement against the existence of supraspinatus calcific tendinopathy in the models.

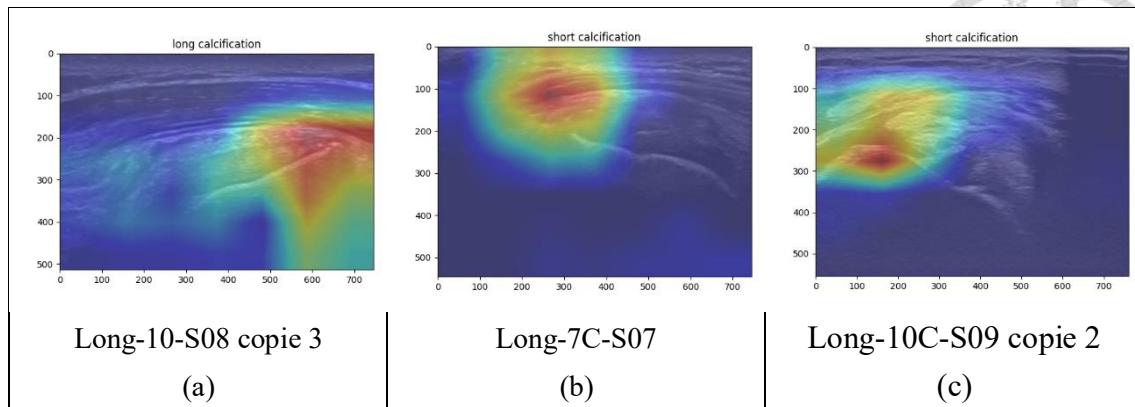
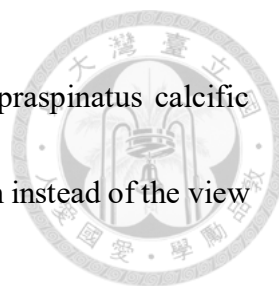


Figure 5-6: Heatmaps with color distributed in greater tuberosity. (a) A heatmap of the longitudinal model. (b) A heatmap of the transverse model. (c) A heatmap of the longi-trans model.

The common points of these sorts are their different levels of correlation with supraspinatus tendon. Heatmaps with color distributed on supraspinatus tendon have no doubt to be the basis to inference that supraspinatus tendon is essential for the diagnosis of supraspinatus calcific tendinopathy. For heatmaps with color distributed under supraspinatus tendon, it is also speculated that not only the supraspinatus tendon itself, but its surrounding area would be the clue of the existence of supraspinatus calcific tendinopathy. As for heatmaps with color distributed in greater tuberosity, it illustrates that surrounding tissues, such as greater tuberosity is also a basis for diagnosis against the existence of supraspinatus calcific tendinopathy.

Heatmaps of models testing against different view of supraspinatus tendon are additional evidences for the ability of longi-trans model. Figure 5-7 is the heatmaps of longitudinal model testing against longitudinal view and transverse view of supraspinatus



tendon. It was found that the model diagnosed the existence of supraspinatus calcific tendinopathy by the calcification occurred in the supraspinatus tendon instead of the view of the supraspinatus tendon.

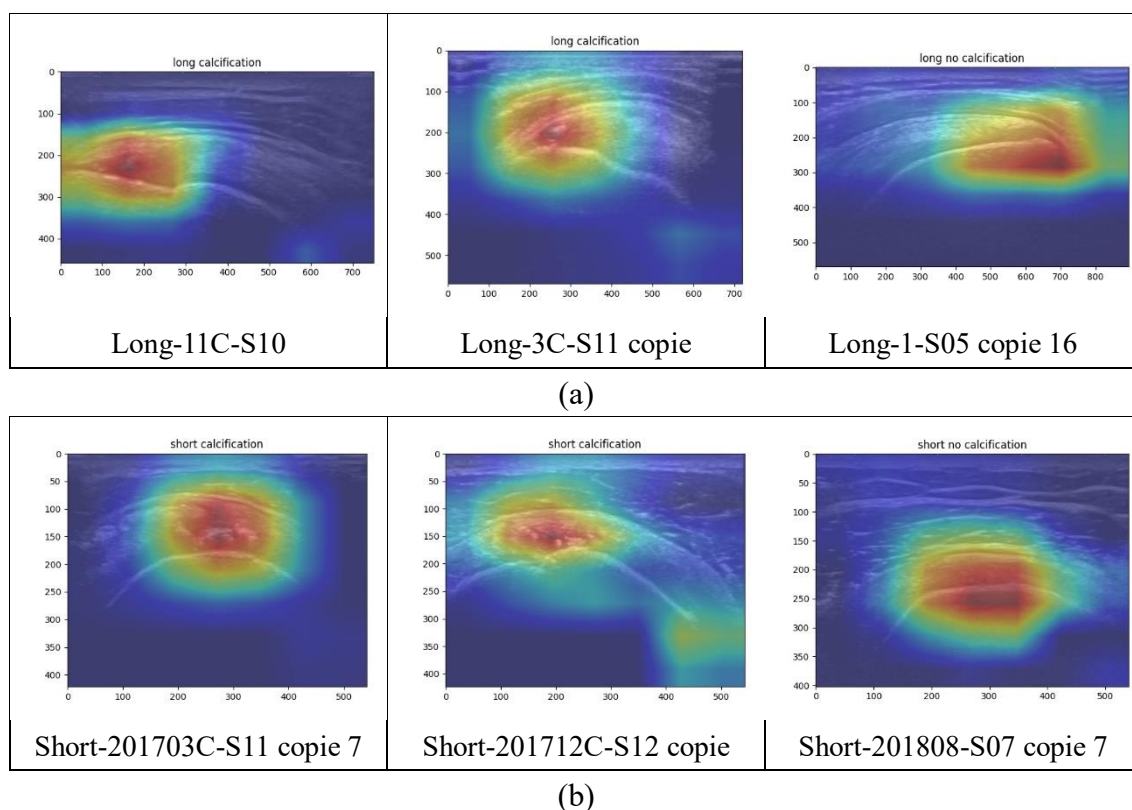
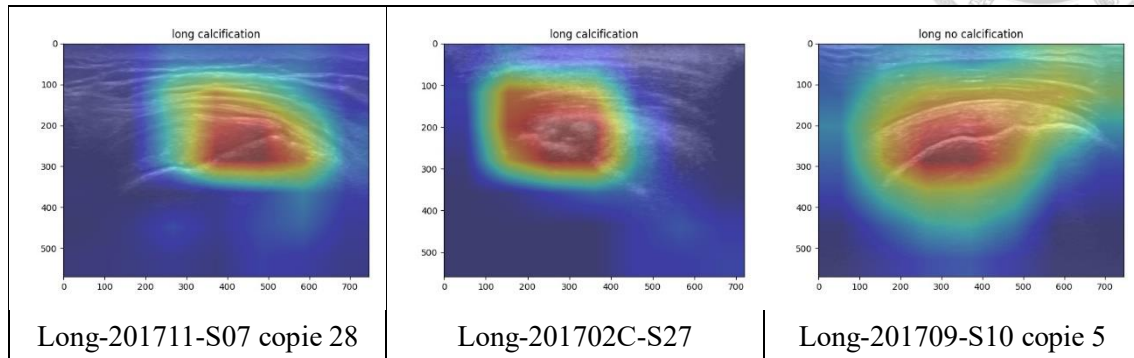


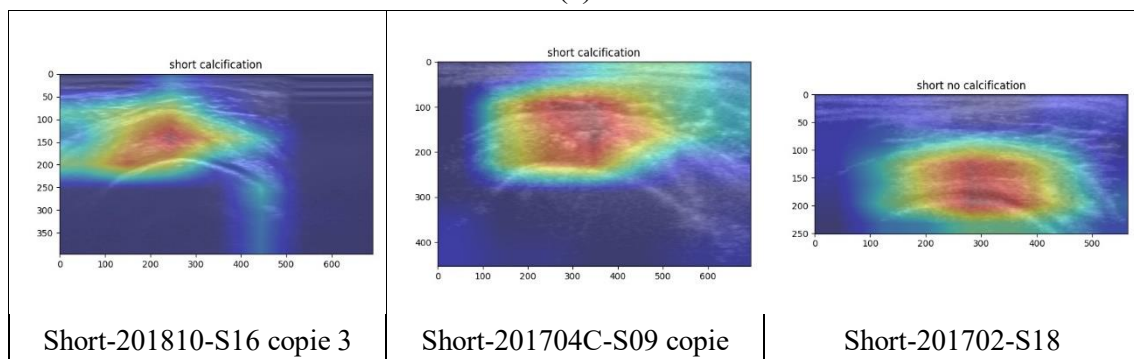
Figure 5-7: Heatmaps of longitudinal model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.

Figure 5-8 and Figure 5-9 are the heatmaps of transverse model and longi-trans model testing against longitudinal view and transverse view of supraspinatus tendon. These heatmaps also shows that models trained on only transverse view of supraspinatus tendon and both longitudinal and transverse view of supraspinatus tendon diagnosed the

existence of supraspinatus calcific tendinopathy by the calcification occurred in the supraspinatus tendon instead of the view of the supraspinatus tendon.

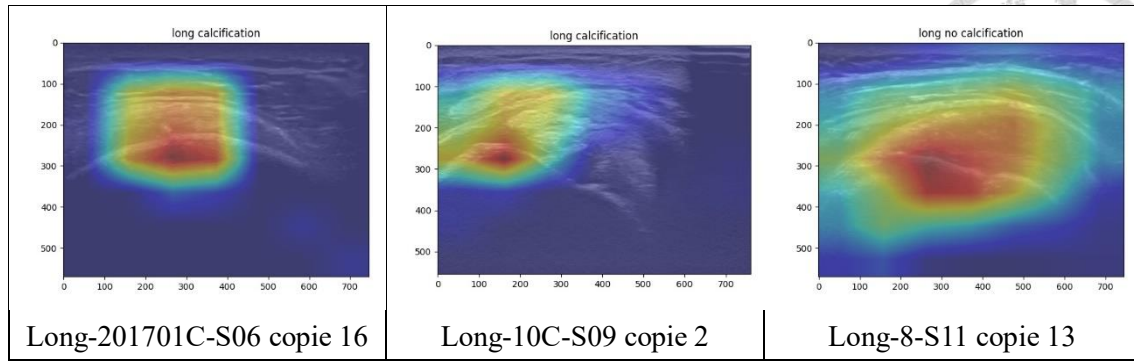


(a)

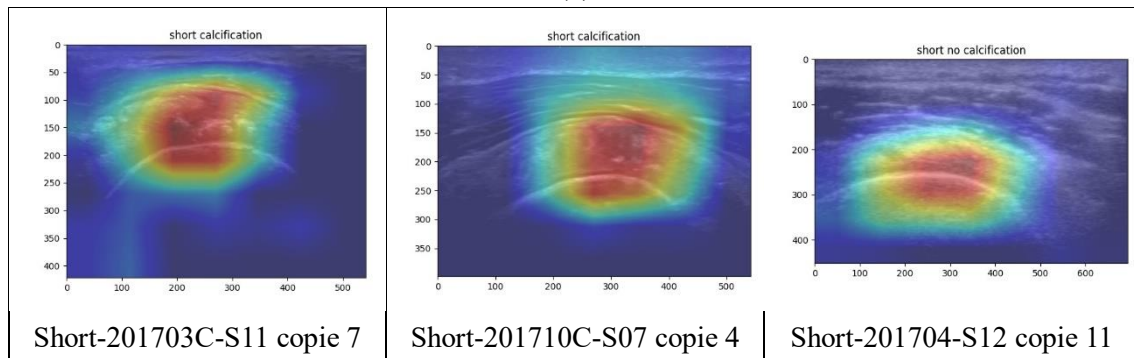


(b)

Figure 5-8: Heatmaps of transverse model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.



(a)



(b)

Figure 5-9: Heatmaps of longi-trans model testing against (a) longitudinal view and (b) transverse view of supraspinatus tendon.

Chapter 6 Conclusion



Supraspinatus calcific tendinopathy accounts for a significant proportion of the US-diagnosed disorder among shoulder pain. Although US has been regarded as an effective way to diagnose supraspinatus calcific tendinopathy, there are still some significant shortages of it. According to the estimation by the World Health Organization, two thirds of the global population lack access to radiology diagnostics (Mollura et al. [47], 2010). Even when radiology diagnostics is available, the operator-dependent characteristic of US makes it difficult to diagnose supraspinatus calcific tendinopathy with high diagnostic accuracy among different experienced level of physicians. Moreover, the uneven workload distribution causes the overwork of physicians which also reduce their diagnosis accuracy against the diagnosis of supraspinatus calcific tendinopathy.

The longi-trans model developed an CNN-based deep learning algorithm in this model is effective on providing physicians professional advice and popularize the radiology diagnostics services by judging the existence of supraspinatus calcific tendinopathy in both longitudinal view and transverse view. This model is more knowledgeable than the longitudinal model and the transverse model with its better performance which is also closer to the actual usage of physicians.

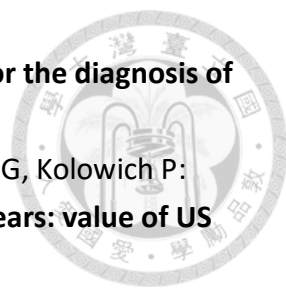
We hope this machine learning application can eliminate the differences of the diagnostic accuracy among different experienced level of physicians, reduce the workload of specialists in radiology, and enhance the accessibility of radiology diagnostics among the global population.

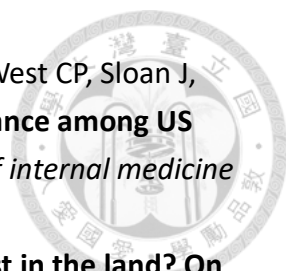




Reference

1. Feleus A, Bierma-Zeinstra S, Miedema H, Bernsen R, Verhaar J, Koes B: **Incidence of non-traumatic complaints of arm, neck and shoulder in general practice.** *Manual Therapy* 2008, **13**(5):426-433.
2. Ottenheijm RP, van't Klooster IG, Starmans LM, Vanderdood K, de Bie RA, Dinant G-J, Cals JW: **Ultrasound-diagnosed disorders in shoulder patients in daily general practice: a retrospective observational study.** *BMC family practice* 2014, **15**(1):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.** *Journal of shoulder and elbow surgery* 2015, **24**(10):1588-1593.
4. Serafini G, Sconfienza LM, Lacelli F, Silvestri E, Aliprandi A, Sardanelli F: **Rotator cuff calcific tendonitis: short-term and 10-year outcomes after two-needle US-guided percutaneous treatment—nonrandomized controlled trial.** *Radiology* 2009, **252**(1):157-164.
5. Gimblett PA, Saville J, Ebrall P: **A conservative management protocol for calcific tendinitis of the shoulder.** *Journal of manipulative and physiological therapeutics* 1999, **22**(9):622-627.
6. Fusaro I, Orsini S, Diani S, Saffioti G, Zaccarelli L, Galletti S: **Functional results in calcific tendinitis of the shoulder treated with rehabilitation after ultrasonic-guided approach.** *Musculoskeletal surgery* 2011, **95**(1):31-36.
7. Farin PU, Jaroma H: **Sonographic findings of rotator cuff calcifications.** *Journal of ultrasound in medicine* 1995, **14**(1):7-14.
8. 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**(1):e23-e23.
9. Churchill SR, Fehringier EV, Dubinsky TJ, Matsen III FA: **Rotator cuff ultrasonography: diagnostic capabilities.** *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 2004, **12**(1):6-11.
10. Al-Shawi A, Badge R, Bunker T: **The detection of full thickness rotator cuff tears using ultrasound.** *The Journal of bone and joint surgery British volume* 2008, **90**(7):889-892.
11. De Jesus JO, Parker L, Frangos AJ, Nazarian LN: **Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis.** *American Journal of Roentgenology* 2009, **192**(6):1701-1707.
12. Iannotti JP, Ciccone J, Buss DD, Visotsky JL, Mascha E, Cotman K, Rawool NM:

- 
- Accuracy of office-based ultrasonography of the shoulder for the diagnosis of rotator cuff tears.** *JBJS* 2005, **87**(6):1305-1311.
13. Jacobson JA, Lancaster S, Prasad A, van Holsbeeck MT, Craig JG, Kolowich P: **Full-thickness and partial-thickness supraspinatus tendon tears: value of US signs in diagnosis.** *Radiology* 2004, **230**(1):234-242.
 14. Rutten MJ, Jager GJ, Kiemeney LA: **Ultrasound detection of rotator cuff tears: observer agreement related to increasing experience.** *American Journal of Roentgenology* 2010, **195**(6):W440-W446.
 15. Sansone V, Consonni O, Maiorano E, Meroni R, Goddi A: **Calcific tendinopathy of the rotator cuff: the correlation between pain and imaging features in symptomatic and asymptomatic female shoulders.** *Skeletal radiology* 2016, **45**(1):49-55.
 16. 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**(3):158-162.
 17. McCormack RA, Nayyar S, Jazrawi L: **Physician training: ultrasound and accuracy of diagnosis in rotator cuff tears.** *Bulletin of the NYU Hospital for Joint Diseases* 2016, **74**(3):207.
 18. O'Connor PJ, Rankine J, Gibbon W, Richardson A, Winter F, Miller JH: **Interobserver variation in sonography of the painful shoulder.** *Journal of Clinical Ultrasound* 2005, **33**(2):53-56.
 19. Ohrndorf S, Naumann L, Grundey J, Scheel T, Scheel AK, Werner C, Backhaus M: **Is musculoskeletal ultrasonography an operator-dependent method or a fast and reliably teachable diagnostic tool? Interreader agreements of three ultrasonographers with different training levels.** *International journal of rheumatology* 2010, **2010**.
 20. Radiology ACo: **ACR practice guideline for performing and interpreting diagnostic ultrasound examinations.** *American College of Radiology* 2006.
 21. Hertzberg BS, Kliewer MA, Bowie JD, Carroll BA, DeLong DH, Gray L, Nelson RC: **Physician training requirements in sonography: how many cases are needed for competence?** *American Journal of Roentgenology* 2000, **174**(5):1221-1227.
 22. Hawkins J: **Addressing the shortage of radiologists.** *Radiology Management* 2001, **23**(4):26-29.
 23. Yawn B, Krein S, Christianson J, Hartley D, Moscovice I: **Rural radiology: who is producing images and who is reading them?** *The Journal of Rural Health* 1997, **13**(2):136-144.
 24. Soni K: **Has the Radiologist Shortage Been Resolved? Recent Findings Using an Improved Survey Based Measurement.** 2009.

- 
25. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, West CP, Sloan J, Oreskovich MR: **Burnout and satisfaction with work-life balance among US physicians relative to the general US population.** *Archives of internal medicine* 2012, **172**(18):1377-1385.
 26. Kaplan A, Haenlein M: **Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence.** *Business Horizons* 2019, **62**(1):15-25.
 27. Mitchell TM: **Machine learning.** 1997. Burr Ridge, IL: McGraw Hill 1997, **45**(37):870-877.
 28. Russell S, Norvig P: **Artificial intelligence: a modern approach.** 2002.
 29. Deng L, Yu D: **Deep learning: methods and applications.** *Foundations and trends in signal processing* 2014, **7**(3-4):197-387.
 30. Bengio Y, Courville A, Vincent P: **Representation learning: A review and new perspectives.** *IEEE transactions on pattern analysis and machine intelligence* 2013, **35**(8):1798-1828.
 31. Schmidhuber J: **Deep learning in neural networks: An overview.** *Neural networks* 2015, **61**:85-117.
 32. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *nature* 2015, **521**(7553):436-444.
 33. Carin L, Pencina MJ: **On deep learning for medical image analysis.** *Jama* 2018, **320**(11):1192-1193.
 34. Hubel DH, Wiesel TN: **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.** *The Journal of physiology* 1962, **160**(1):106.
 35. LeCun Y, Bottou L, Bengio Y, Haffner P: **Gradient-based learning applied to document recognition.** *Proceedings of the IEEE* 1998, **86**(11):2278-2324.
 36. Krizhevsky A, Sutskever I, Hinton GE: **Imagenet classification with deep convolutional neural networks.** In: *Advances in neural information processing systems: 2012*; 2012: 1097-1105.
 37. Simonyan K, Zisserman A: **Very deep convolutional networks for large-scale image recognition.** *arXiv preprint arXiv:14091556* 2014.
 38. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A: **Going deeper with convolutions.** In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2015*; 2015: 1-9.
 39. He K, Zhang X, Ren S, Sun J: **Deep residual learning for image recognition.** In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2016*; 2016: 770-778.
 40. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ: **Densely connected convolutional networks.** In: *Proceedings of the IEEE conference on computer*

- 
- vision and pattern recognition: 2017; 2017: 4700-4708.*
41. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE: **Machine learning for medical ultrasound: status, methods, and future opportunities.** *Abdominal radiology* 2018, **43**(4):786-799.
 42. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR: **Improving neural networks by preventing co-adaptation of feature detectors.** *arXiv preprint arXiv:12070580* 2012.
 43. West JV, Dan; Warnick, Sean: **"Spring Research Presentation: A Theoretical Foundation for Inductive Transfer"**. *Brigham Young University, College of Physical and Mathematical Sciences* 2007.
 44. Loua T: **Atlas statistique de la population de Paris:** J. Dejeu & cie; 1873.
 45. 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: 2017;* 2017: 618-626.
 46. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A: **Learning deep features for discriminative localization.** In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2016; 2016: 2921-2929.*
 47. Mollura DJ, Azene EM, Starikovskiy A, Thelwell A, Iosifescu S, Kimble C, Polin A, Garra BS, DeStigter KK, Short B: **White paper report of the RAD-AID Conference on International Radiology for Developing Countries: identifying challenges, opportunities, and strategies for imaging services in the developing world.** *Journal of the American College of Radiology* 2010, **7**(7):495-500.